THE

LONDON, EDINBURGH AND DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE

CONDUCTED BY

SIR OLIVER JOSEPH LODGE, D.Sc., LL.D., F.R.S.
SIR JOSEPH JOHN THOMSON, O.M., M.A., Sc.D., LL.D., F.R.S.
AND
ALLAN FERGUSON, M.A., D.Sc.

"Nec aranearum sane textus ideo melior quia ex se fila gignunt, nec noster vilior quia ex alienis libamus ut apes."

Just. Lips. *Polit.* lib. i. cap. 1. Not.

VOL. XXIX.—SEVENTH SERIES

JANUARY—JUNE 1940

LONDON:

TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET STREET, E.C. 4.
Agents: Wm. DAWSON & SONS, LTD., CANNON HOUSE, PILGRIM STREET, E.C. 4.

"Meditationis est perscrutari occulta; contemplationis est admirari perspicua . . . . Admiratio generat quæstionem, quæstio investigationem, investigatio inventionem."—*Hugo de S. Victore.*

—— "Cur spirent venti, cur terra dehiscat,
Cur mare turgescat, pelago cur tantus amaror,
Cur caput obscura Phœbus ferrugine condat,
Quid toties diros cogat flagrare cometas,
Quid pariat nubes, veniant cur fulmina cœlo,
Quo micet igne Iris, superos quis conciat orbes
Tam vario motu."

J. B. Pinelli ad Mazonium.

ALERE FLAMMAM.

CONTENTS OF VOL. XXIX. (Nos. 192–197)

(SEVENTH SERIES)

NUMBER 192—JANUARY

Page

I. The Transition from Glow to Arc Discharge. By E. L. E. Wheatcroft, M.A., and H. Barker, M.Sc. (Plates I. & II.) 1

II. A Note on the Schottky Effect. By E. L. E. Wheatcroft 16

III. Magnetic Screening by Hollow Circular Cylinders. By J. C. Jaeger, Lecturer in Mathematics in the University of Tasmania 18

IV. On Iso-dielectric Lines and Geologic Structure. By H. Löwy, Dr.phil. 32

V. The Electric Field produced by a Point-charge located outside a Dielectric Wedge. By S. O. Rice 36

VI. A Mechanical Analogy to the Motion of Electrons in Gases. By G. D. Yarnold, University College, Nottingham 47

VII. The Radiation Correction in Calorimetry. By F. E. Hoare, Ph.D., A.R.C.S. 52

VIII. The Method of Symmetrical Components applied to Harmonic Analysis. By Louis A. Pipes, Ph.D., The Harvard Engineering School, Cambridge, Mass. 66

IX. On a Hystero-Differential Equation arising in a Probability Problem. By Ludwik Silberstein, Ph.D. 75

X. Effusion Phenomena in Relativistic Quantum Statistics. By D. V. Gogate, M.Sc., Ph.D. (London), Physics Department, Baroda College, Baroda (India) 85

XI. On some Problems in the Diffraction of Heat. By Dr. Arnold N. Lowan 93

XII. Error in Paper on "Certain Vibration Problems solved by means of an Analogous Problem in Heat Conduction," Phil. Mag. xxii. p. 1079. By G. Green 100

XIII. Notices respecting New Books :—
  Prof. A. S. Eve's Rutherford : being the Life and Letters of the Rt. Hon. Lord Rutherford, O.M. 103
  Mr. H. Semat's Introduction to Atomic Physics 103
  Prof. S. Chapman and Mr. T. G. Gowling's The Mathematical Theory of Non-Uniform Gases 104

MG424
75

NUMBER 193—FEBRUARY

Page

XIV. On the Mutual Electrical Energy of Two Colloidal Particles: General Theory, using Approximate Debye-Hückel Equation. By S. Levine, Department of Colloid Science, Cambridge University, and G. P. Dube, Institut Henri Poincaré, Paris 105

XV. Thermionic Emission, Migration, and Evaporation of Barium on Tungsten. By J. A. Becker, Ph.D., and G. E. Moore, Bell Telephone Laboratories, New York, N.Y. (Plates III.–V.) 129

XVI. The Radiation of the Acoustic Air-Jet Generator derived from direct Observation of the Amplitude of the Aerial Vibrations in the Oscillator. By Jul. Hartmann and Freimut Lazarus. (Plates VI.–VIII.) 140

XVII. On the Theory of Light-scattering. By S. Parthasarathy, D.Sc., F.Inst.P. (Nobel Institute, Stockholm) 148

XVIII. The Musical Scale of the Highland Bagpipe. By G. E. Allan, D.Sc., Applied Physics Department, Glasgow University 154

XIX. A Note on the Carey-Foster Method of Measuring a Mutual Inductance in Terms of a Capacity. By E. Tyler, D.Sc., M.Sc., F.Inst.P., A.F.R.Ae.S., Science Department, South-East Essex Technical College, Dagenham 162

XX. Interaction between Radiation and Matter. By Prof. K. C. Kar, D.Sc. 169

XXI. The Specific Heat of Water and of Heavy Water. By A. H. Cockett and Allan Ferguson 185

XXII. On the Proton-Proton Scattering. By Prof. K. C. Kar, D.Sc., and D. Basu, B.Sc. 200

XXIII. Secrecy by Inversion of Syllables. By Mario Marro, Dr. Ing. 205

XXIV. Notices respecting New Books:—

Mr. James H. Hibben's The Raman Effect and its Chemical Applications 208

NUMBER 194—MARCH

XXV. Stability and Configuration of the Wakes produced by Solid Bodies moving through Fluids. By S. H. Hollingdale, M.A., Ph.D., D.I.C. (Plates IX.–XIII.) 209

XXVI. The Principles of a Mechanical Method for Calculating Regression Equations and Multiple Correlation Coefficients and for the Solution of Simultaneous Linear Equations. By T. E. W. Schumann, M.Sc., Ph.D. 258

XXVII. Note on the Distributions of Temperature and Vapour Pressure around a Horizontal Wet Cylinder. By R. W. Powell, Ph.D., Physics Department, The National Physical Laboratory, Teddington, Middlesex .. 274

XXVIII. A "Scale of Three" Counter. By P. Ohlin, Fil.Lic. 285

XXIX. On the Proposition that "Shearing Stress alone is effective in producing Elastic Failure." By James J. Guest 289

XXX. On the Interrelation of Molecular Constants for Diatomic Molecules.—II. By Dr. R. A. Newing, University College of North Wales, Bangor 298

Page
XXXI. The Effect of a Magnetic Force on High Frequency Discharges in Pure Gases. By A. E. Brown, B.A., B.Sc., Merton College, Oxford .. 302
XXXII. Notices respectin New Books :—
Professor H. C. Plummer's Probability and Frequency 310
Dr. A. W. Barton's A Text Book on Light 311
Mr. J. S. Rogers's Physics for Medical Students 311
Mr. M. H. A. Newman's Elements of the Topology of Plane Sets of Points 312

NUMBER 195—APRIL

XXXIII. The Propagation of Electromagnetic Waves in an Atmosphere containing Free Electrons. By L. G. H. Huxley, M.A., D.Phil., F.Inst.P., F. W. Bennett Lecturer in Physics, University College, Leicester 313
XXXIV. The Theory of the Electric Charge and the Quantum Theory.—Part I. By H. T. Flint, Reader in Physics, University of London, King's College 330
XXXV. Dielectric Constant of Ionized Air.—III. By Shah Md. Fazlur Rahman, M.Sc., and S. R. Khastgir, D.Sc.(Edin.), Mem.I.R.E., Dacca University 344
XXXVI. The Dielectric Constant and Electrical Conductivity of Gases and Vapours Ionized by X-rays at Ultra-high Radio Frequency. By S. R. Khastgir, D.Sc.(Edin.), Mem.I.R.E., and Shah Md. Fazlur Rahman, M.Sc., Dacca University 353
XXXVII. The Temperature Variation of Viscosity in Non-Polar and Dipolar Inorganic Liquids. By S. T. Bowden, D.Sc., and A. R. Morgan, M.Sc., University College, Cardiff 367
XXXVIII. The Rôle of Ionization by Positive Ions in Spark Breakdown. By R. N. Varney, Ph.D., Assistant Professor of Physics, Washington University, St. Louis, Mo., L. B. Loeb, Ph.D., Professor of Physics, University of California, Berkeley, Calif., and W. R. Haseltine, Ph.D., Research Associate in Physics, University of California, Berkeley, Calif... 379
XXXIX. Experiments on the Action of Ionizing Radiations on the Electrophoretic Mobility of Colloid Particles. By J. A. Crowther, M.A., Sc.D., Professor of Physics in the University of Reading, H. Liebmann, Ph.D., and R. Jones, B.Sc. 391
XL. Observations on Edge Tones. By J. M. A. Lenihan, B.Sc., and E. G. Richardson, B.A., Ph.D., D.Sc. (University of Durham, King's College, Newcastle-upon-Tyne) 400
XLI. Waves produced by the Rolling of a Ship. By T. H. Havelock, F.R.S. 407
XLII. Notices respecting New Books :—
Terrestrial Magnetism and Electricity 415
Dr. Arthur I. Vogel's A Text-Book of Quantitative Inorganic Analysis Theory and Practice 416

NUMBER 196—MAY

Page

XLIII. The Theory of the Electric Charge and the Quantum Theory.—Part II. By H. T. Flint, Reader in Physics, University of London, King's College 417

XLIV. A Rigorous Theory of the Recombination of both small and large Ions in Gases at High Pressures. By W. R. Harper, Ph.D., Research Associate, H. H. Wills Physical Laboratory, University of Bristol........ 434

XLV. The Fundamental Principles of Energetics. By J. N. Brønsted (Institute of Physical Chemistry, the University of Copenhagen)........ 449

XLVI. Gravity and the Rotation of the Earth. By C. G. Pendse, M.A., Ph.D. 471

XLVII. On Mass and Force in Newtonian Mechanics—Addendum to "Mass I." and "Mass II." By C. G. Pendse, M.A., Ph.D. 477

XLVIII. Iron-cored Coils of High Efficiency. By G. F. Partridge, B.Sc., F.Inst.P. 485

XLIX. The Equations of Motion of Point Electrons deduced from a Variational Principle. By T. Lewis, Aberystwyth 495

L. Haidinger's Rings in Non-Uniform Plates. By Sir C. V. Raman and V. S. Rajagopalan. (Plates XIV. & XV.) 508

LI. Note on "The Theory of Light-scattering." By R. S. Krishnan.. 515

LII. Notices respecting New Books :—

Prof. Erwin Meyer's Electro-acoustics 517

Prof. J. Reilly and Prof. W. N. Rae's Physico-chemical Methods. 517

Sir A. S. Eddington's The Philosophy of Physical Science...... 518

Mr. Duncan A. MacInnes's The Principles of Electrochemistry.. 519

Mr. N. A. McKenna's Theoretical Electrochemistry 520

NUMBER 197—JUNE

LIII. Theory of Resonance in Microwave Transmission Lines with discontinuous Dielectric. By H. R. L. Lamont, M.A., B.Sc., Carnegie and Breadalbane Scholar in the University of Glasgow 521

LIV. The Stresses in a Circular Ring supported on Frictionless Radial Pins. By Professor A. J. Sutton Pippard, M.B.E., D.Sc., Imperial College, London 541

LV. On Flint's Five-Dimensional Theory of the Electron. By William Band, Professor of Physics, Yenching University, Peking, China 548

LVI. The Structure of Meteoritic Iron. By Professor E. A. Owen, M.A., Sc.D., University College of North Wales, Bangor. (Plate XVI.)........ 553

LVII. Note on a Method for the Comprehensive Analysis of Complex Power Plant. By A. C. Walshaw, M.Sc., D.I.C. (Lecturer in Applied Thermodynamics at the Northampton Polytechnic, London) 568

LVIII. Some Impedance Characteristics of Tapped Resonant-line Circuits. By W. E. Willshaw, M.Sc.Tech. (Communication from the Staff of the Research Laboratories of The General Electric Company Limited, Wembley, England.) 572

LIX. Differential Operators acting as Integrators. By Ludwik Silberstein, Ph.D. 586

Page
LX. Note on Paper on "Calculation of Triode Constants." By S. Rodda 603
LXI. Erratum to the Paper "On the Interrelation of Molecular Constants for Diatomic Molecules.—II." By R. A. Newing 603
LXII. Notices respecting New Books :—
Mr. C. J. Coe's Theoretical Mechanics 604
The Mathematical Papers of Sir Wm. Rowan Hamilton.—Vol. II. Dynamics. Edited by Professors Conway and McConnell 604
Prof. Harnwell's Principles of Electricity and Magnetism 605
Mr. C. B. Boyer's The Concepts of the Calculus 605
Prof. H. Levy's Modern Science : a Study of Physical Science in the World Today 606
Mr. Rollo Appleyard's The History of the Institution of Electrical Engineers, 1871–1931 606
Mr. T. L. Kelley's The Kelley Statistical Tables 607
Mr. W. B. Fite's Advanced Calculus 607
Mr. N. W. McLachlan's Complex Variable and Operational Calculus 607
Mr. F. Bowman's Introduction to Bessel Functions 608

Index 609

PLATES

I. & II. Illustrative of Prof. E. L. E. Wheatcroft and Mr. H. Barker's Paper on the Transition from Glow to the Discharge.
III.–V. Illustrative of Dr. J. A. Becker and Mr. G. E. Moore's Paper on Thermionic Emission, Migration, and Evaporation of Barium on Tungsten.
VI.–VIII. Illustrative of Messrs. Jul. Hartmann and Freimut Lazarus's Paper on the Radiation of the Acoustic Air-Jet Generator derived from direct Observation of the Amplitude of the Aerial Vibrations in the Oscillator.
IX.–XIII. Illustrative of Dr. S. H. Hollingdale's Paper on Stability and Configuration of the Wakes produced by Solid Bodies moving through Fluids.
XIV. & XV. Illustrative of Sir C. V. Raman and Mr. V. S. Rajagopalan's Paper on Haidinger's Rings in Non-Uniform Plates.
XVI. Illustrative of Prof. E. A. Owen's Paper on the Structure of Meteoritic Iron.

THE

LONDON, EDINBURGH & DUBLIN

PHILOSOPHICAL MAGAZINE

AND

JOURNAL OF SCIENCE

[SEVENTH SERIES—VOL. 29]

I. *The Transition from Glow to Arc Discharge.*

By E. L. E. Wheatcroft, M.A., and H. Barker, M.Sc.*

[Received January 20, 1939.]

[Plates I. & II.]

It has been known from the early days of vacuum technique that at low pressure the electrical breakdown of a gas yields what is usually called a glow, while at higher pressures the observed effect is a spark. In an earlier paper † we have recorded measurements of the boundary zone between these two phenomena, showing that, quite apart from the natural random spread, the spark is developed somewhat gradually with increasing pressure, in that there is an increasing concentration of the luminosity into a bright central core, which luminosity eventually breaks through the Faraday dark space to reach the cathode. Our records were, of course, only of visual or photographic impressions; Rogowski and Tamm ‡, however, who had the advantages of an impulse generator and cathode-ray oscillograph, were able to show, at any rate under the conditions of their experiments, that the spark was only developed as a second stage of breakdown subsequent to the glow. According to their results the time interval between first and second breakdown increased from a negligibly small value at atmospheric pressure to a value which became infinite, and so yielded a stable glow, at around 7 mm. Hg. in air. We have always regarded this as rather important evidence in explanation of what at first sight would be an unexpected effect, namely, that Paschen's law holds equally over the range of sparks and glows *; unfortunately, however, the results do not seem to have been repeated by other experimenters.

---

* Communicated by the Authors.

† Phil. Mag. xx. (1935) p. 562. It seems incredible that these should be the first measurements of the boundary zone, but we have not been able to find any earlier records.

‡ *Arch. f. Elekt.* xx. p. 625 (1928).

Rogowski and Tamm found that the glow is stable at constant voltage below a pressure of some 7 mm. Hg. The existence of a stable glow at low pressure, with its normal or abnormal cathode fall, is, of course, not questioned; it is known too that if the applied voltage is raised after a glow is formed there is eventually reached a point where the glow breaks down to an arc. Wehrli † made extensive measurements with tungsten electrodes in nitrogen, finding apparently a smooth transition from glow to arc. More recently von Issendorf, Schenkel, and Seeliger ‡, looking for causes of backfiring in mercury arcs, have drawn attention to "scintillations" observed on a glow cathode when the glow voltage becomes too high. This suggests that before any arc is established there occurs intermittent breakdown, in the sense of sudden collapse and recovery of voltage similar to the well-known relaxation of an intermittent spark.

In the experiments described here we have recorded with a cathode-ray oscillograph the initial stages of transition from the abnormal glow for copper electrodes in air at pressures from 0·2 to 1·2 mm. Hg. As may be seen from fig. 3 (Pl. I.) we obtained direct evidence of intermittent voltage collapse, while we were able to avoid any protracted arcing with its attendant experimental troubles.

1. *Experimental Arrangement.*

Fig. 1 shows the essentials of the circuit employed. Previous experience had shown that the energy which flows from a generator into an arc even in the short time of tripping a circuit breaker is so great that reproducible results could not be obtained. The "source" was therefore a bank of condensers of 4 microfarads total capacity, charged to 12,000 volts. In this way the energy of the discharge was limited to a value which allowed several score of "shots" to be made before the tube had to be dismantled and new electrodes fitted.

---

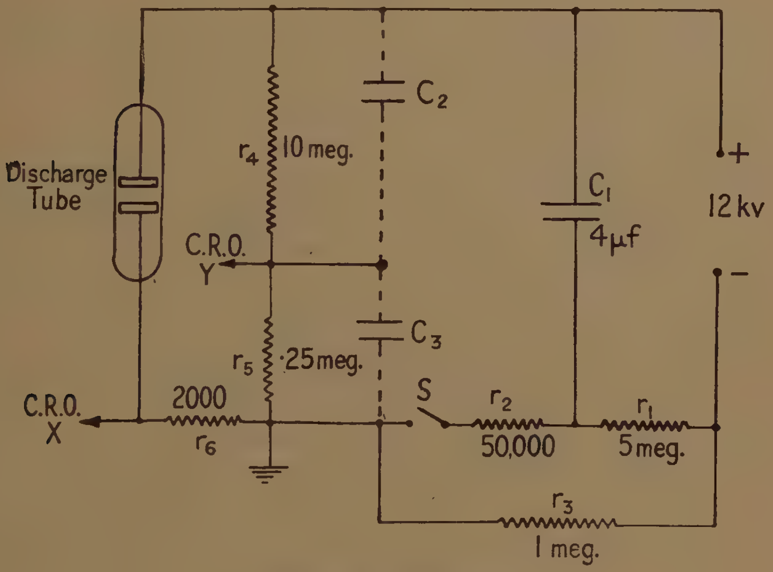
* Phil. Mag. xx. p. 571 (1935).

† *Zeits. f. Phys.* xliv. p. 301 (1927).

‡ *Wiss. Veroff. aus. d. Siemenskonzern.* ix. p. 73 (1930).

The bank of condensers C_1 was charged through the high resistance r_1 to the voltage of the supply. On closing the switch S the condensers then discharged through the resistances r_2, r_6 and the tube, the maximum current being controlled by the resistance r_2. The rate of rise of voltage across the discharge-tube was limited only by the self-capacity of the electrodes and wiring of the circuit; since these capacities are small the time of rise was less than 10^{-4} seconds, except where special arrangements were made to delay it.

Fig. 1.



Electrical connexions.

It was found necessary to maintain a small initial discharge current through the tube by means of a high-resistance r_3; for if the rising voltage is applied to the discharge-tube without the initial current, the glow does not form until the voltage has reached a value at which the glow current is comparatively high. In that case the current, instead of rising, starts at a high value and begins to fall immediately. When, on the other hand, a small initial glow current is used, the current starts at this value and rises with the voltage until the transition sets in or until both fall again due to the discharge of the condenser.

The voltage across the discharge was indicated by connecting across the tube a potential divider $r_4\ r_5$ having a resistance ratio of 40 : 1. The junction between the two portons was connected to one " Y " deflexion plate of the oscillograph, the other " Y " plate being earthed except where disconnected, as described later, for a special experiment. The potential divider was compensated to preserve its ratio at all frequencies by connecting a variable air condenser C_2 across the high-resistance section of the divider. The capacity C_3 is provided by the copper screening boxes described below. The condenser C_2 was adjusted so that the ratio and phase angle of the potential divider remained the same at frequencies 50, 10,000, and 10^6 cycles/sec.

The current to the discharge-tube was indicated on the oscillograph by developing a voltage across a resistance r_6, and applying it to one " X " deflexion plate of the oscillograph, the other " X " plate being earthed. It is, of course, an inherent difficulty that, owing to the self-capacity of the discharge-tube, it is impossible to record directly the actual ionic current during any rapid change of tube potential.

The potential divider, the current shunt, and the leads from them to the oscillograph were enclosed in stout copper screening boxes.

The oscillograph was a sealed-off hot cathode tube of the familiar television type, with an acceleration of 3000 volts. We experienced considerable experimental difficulties due to radiation from the discharge-tube itself. The development of the arc was always accompanied by defocusing of the beam so that the spot on the fluorescent screen became an irregular splash. The effect could be minimized by connecting several ·0001 mfd. mica condensers between the electron gun electrodes of the oscillograph, the connexions being made at the terminals on the base of the tube, but it was found impossible to remove the effect altogether by screening or any other method. The radiation was clearly of very short wave-length, and developed a considerable potential across a foot length of wire. It was therefore rather a matter of good fortune that the defocusing effect made itself felt only during the later stages of the development of the arc, and so it was quite possible to obtain clear impressions of the initial stages of the breakdown, which were the object of our measurements.

When observing voltage-time and current-time relations on the oscillograph the timing traverse was obtained in the usual way by charging a condenser through a pentode valve. A bias voltage applied to one plate kept the beam off the screen until the traverse began.

In the arrangement finally adopted the synchronization of the circuits was accomplished by a switch (S in fig. 1) which applied the voltage to the discharge while a second switch started the oscillograph traverse by opening the short circuit across the timing condenser, the two switches being coupled rigidly together and operated by a falling weight. By this means the rate of travel of the switch arms was definite and repeatable.

Experiment showed that the transition to the arc was not affected by the rate of rise of voltage and current up to the maximum rate obtainable with the circuit, viz., about 5×10^7 volts per second, and in the experiments described here no attempt was made to delay the rate of voltage rise.

The high-tension supply was taken from a bank of six rectifying valves fed from a three-phase transformer. The output of the plant was 600 mA. at 12,000 volts D.C. with parallel connexion of the valves, and 300 mA. at 24,000 volts D.C. with series connexion of the valves. The low-tension A.C. to the transformer was supplied from a three-phase alternator, the voltage of which could be varied by means of the field rheostat.

A double-pole switch was used to isolate the discharge circuit from the high-tension supply, since it was found that the high-frequency oscillations in the discharge were liable to cause breakdown in the supply plant.

The camera used to photograph the fluorescent screen of the oscillograph had a fixed focus using a single lens of approximate 3 inches focal length and an approximate aperture of $f/2$. Such a lens is liable to give field and achromatic distortion, but the latter defect is more or less absent in this case owing to the fact that the light emitted from the screen is almost monochromatic, and the former can be allowed for.

In the earlier stages of the work oscillograph recorder paper was used in the camera, but a tremendous improvement in speed resulted by the use of Ilford Hypersensitive Panchromatic Plates. It was found possible to photograph the spot as it travelled across the screen in about 200 μsec.*. A transient positive potential was applied to the modulation grid of the oscillograph tube for the short period that the spot was on the screen; this eliminated fogging of the plate at the beginning and end of the picture due to stray fluorescence.

Since the eye is more sensitive and accepts a much bigger range of contrast than the camera, it was possible to supplement the photographic oscillograms with visual observations. In these cases the values could not be measured, as the eye could not accurately estimate deflexions. However, by sketching the observed phenomena on to prints of the photographs much valuable information could be obtained.

The whole of the following investigations were made with discharges between copper electrodes in air at pressures between 0·2 and 1·4 mm. of mercury.

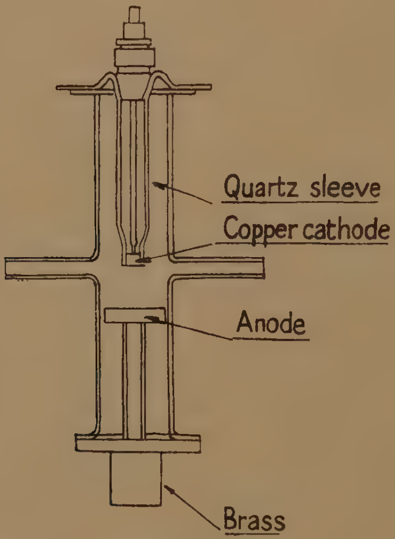
A discharge-tube as shown in fig. 2 was used. This consisted of a cylindrical glass tube with two side arms and ground flanges. The cathode was a small cylinder of copper, 6 mm. diam., surrounded by a quartz sleeve with a flange to fit that of the discharge-tube. Only the end of the electrode was thus left uncovered, giving a very small area of cathode surface, and so enabling the current density to reach high values with low values of current.

---

* Since this time an improvement in this technique is reported by McGillewie (Jl. I. E. E. Nov. 1938).

The anode consisted of another cylinder of copper supported from the brass base of the discharge-tube. The anode entirely filled the bore of the discharge-tube, this being found necessary to prevent the discharge striking across the longer path from the cathode to the brass base instead of the anode.

Fig. 2.



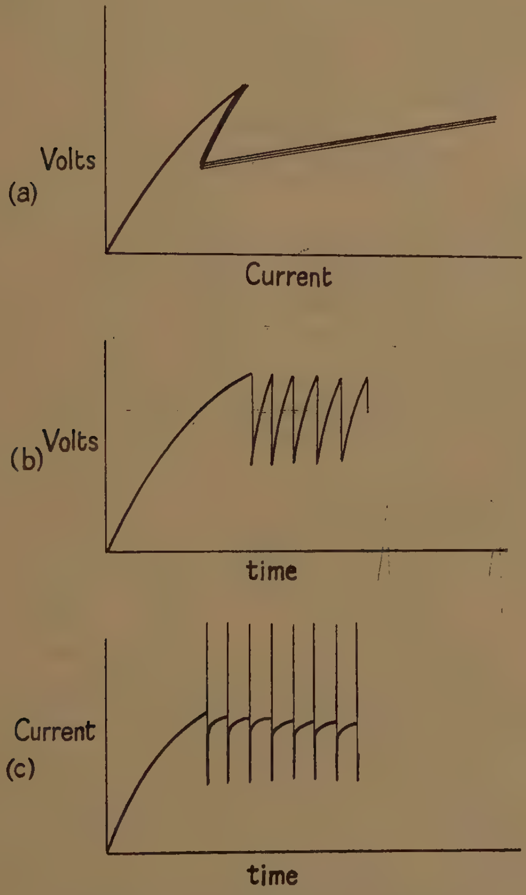
Discharge-tube.

The discharge-tube was exhausted by a rotary oil-pump, one of the side limbs being connected to a short vacuum main having McLeod and Pirani gauges, and a P_2O_5 drying-tube. Dry air was supplied to the system from a bottle containing calcium chloride.

2. *Oscillographic Observations.*

Fig. 3 (Pl. I.) shows typical photographic records of voltage and current of the transition stage. In fig. 3 (*a*) the traverse speed is slow, in order to show the voltage rise and first breakdowns as the crest is reached. This photograph was taken in one of the earlier experiments when the rise was purposely delayed. In figs. 3 (*b*) and (*c*) the traverse speed is increased to demonstrate the wave form of the recovery between the breakdowns. For further resolution a traverse speed too great for photography was needed, and visual observation was used.

Fig. 4.



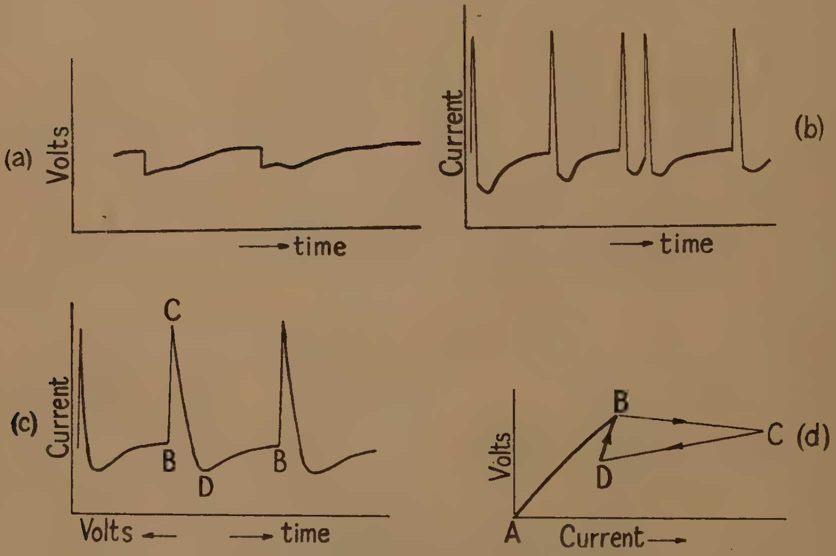
Wave-form of intermittent breakdown.

Figs. 4 (*a*), (*b*), and (*c*) show current, voltage, and time relations as deduced from photographs supplemented by visual observation. Each of the breakdowns is seen to result in typical relaxation oscillations familiar in spark discharges.

It was impossible to obtain a good photograph of the voltage-current characteristic on account of the different intensity of the beam during the stable and oscillatory parts, and this trouble of course also militated against visual observations. What was seen was always of the form of fig. 4 (*a*).

Fig. 5 illustrates the method used to correlate the three types of record. Figs. 5 (*a*) and (*b*) are sketches of visual observations with a rapid traverse. Since the observations were of transient discharges (and therefore fig. 5 (*a*) and (*b*) do not refer to the same " shot ") there is so far no method of correlating the time relations of the voltage and current observations; for in the voltage-current observations in fig. 4 (*a*) a large number of breakdowns are superposed, and thus it was impossible to see clearly there what was happening. In fig. 5 (*c*), therefore, the connexions were similar to those for the current-time relations, except that the *voltage* from the potential divider was simultaneously applied to the remaining " X " plate in such a way as to deflect the beam in the opposite direction to the direction of time traverse. By this method the volt-ampere characteristic of each individual oscillation (where the speed of the beam due to change of current and voltage was high compared with the speed due to the time traverse) could be observed, whilst individual oscillations were separated by the traverse. The beam intensity could then be much greater than for the simple voltage-current record shown in fig. 4. The portions marked BCD in fig. 5 (*c*) were not

Fig. 5.



Interpretation of oscillograms.

always visible on the screen (in some cases there appeared simply a gap from B to D), but they were observed sufficiently often to enable interpretation of the voltage-current characteristic to be made on the lines of fig. 5 (*d*). The rising portion AB corresponds to the abnormal glow characteristic. When breakdown begins the beam shoots very rapidly to C with a voltage slightly below that of B. The current and voltage then fall together to D, with recovery along the line DB. The oscillations thus occur round the triangle BCD in the sense of the arrows of the figure.

3. *Interpretation of Oscillogram.*

The analysis described above shows that the transient breakdowns are initiated as a very rapid rise of current (along BC in fig. 5) with comparatively small drop of voltage across the potential divider. Now the magnitude of the current peak BC, with a shunt resistance r_6 of 2000 ohms, was almost always off the range of the oscillograph, *i. e.*, the instantaneous drop across r_6 was at least several hundred volts. We found, moreover, that there was no apparent reduction in this amplitude when r_6 was changed to 200 ohms. The oscillogram must therefore be interpreted as showing that the breakdowns are initiated by a sudden drop in the inter-electrode potential corresponding to a partial discharge of the self-capacity of the electrodes, and that the magnitude of this drop is not (as far as we could tell) affected by the external circuit. The recovery along CDB is, of course, then easily interpreted as the discharge of C_2C_3 through r_6 and subsequent recharge through r_2.

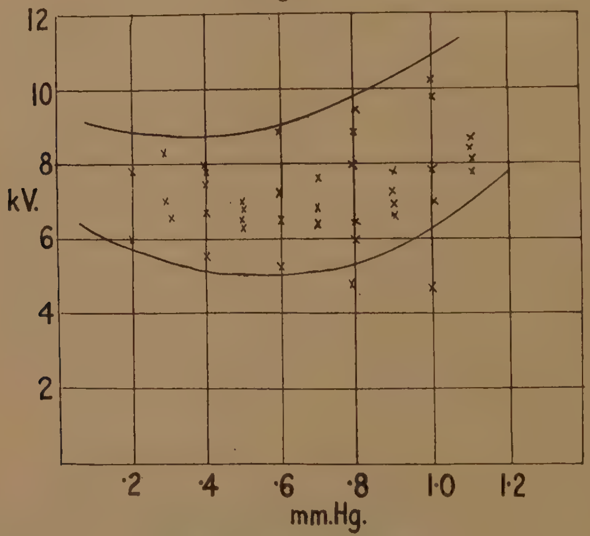
4. *Limits of Glow Current Density.*

The shape of the voltage-current characteristic of the discharge seems to justify the view that the point at which the breakdowns start is the commencement of the transition from glow discharge to arc discharge. Beyond it the discharge can no longer be regarded as a glow. Moreover, during the experiments much cathode disintegration occurred. Fig. 6 (Pl. II.) is a photograph of the cathode after a few hundred discharges: the original surface was highly polished. Copper from the cathode was deposited heavily on the walls of the discharge-tube. During the discharge "scintillations," *i. e.*, small bright flashes, occurred on the cathode at high current densities, and in some case these developed into a hot spot.

The method of finding the critical upper limit of the glow was as follows. Different reservoir condenser voltages (C_1 in fig. 1) were employed in increasing steps until the oscillograph indicated that breakdowns were occurring. This was found by observing either voltage-current, voltage-time, or current-time relationships on the oscillograph screen. The voltage-time relation gave the most useful indications, since the occurrence of a steep drop in the voltage wave (see fig. 3 (*b*)) was more easily seen than the steep increase which occurred in the current-time relation or the change in direction of the volt-ampere characteristic. A critical condenser voltage was noted at which at least one and not more than two breakdowns were observed during a traverse of the screen. With this method there were intervals of steady glow during which the value of the glow current and voltage could be measured on the oscillograph. These values were taken as the critical upper limit.

Preliminary experiments were performed to determine how far the previous history of the gas affected the results, and it was found that most consistent results were obtained if new air was admitted and a discharge just below the critical value passed for 1½ minutes: if this was not done the volt-ampere characteristic of the glow was found to be changing. The question of the nature of the gas was, however, left for subsequent research.

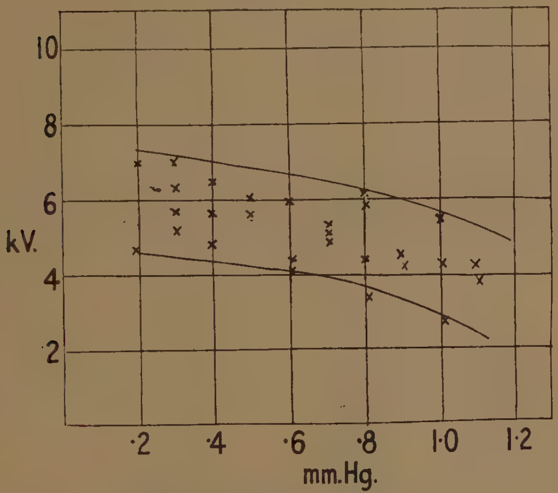
Fig. 7.



Critical condenser voltage.

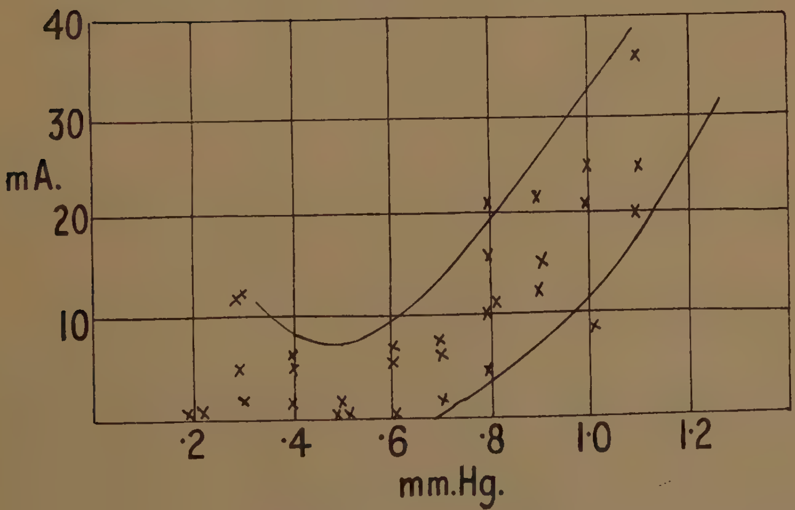
The experiments were carried out over a range of air pressure from 0·2 to 1·1 mm. Hg, and the results are plotted in the three curves figs. 7, 8, and 9, which show variation of critical condenser voltage, corresponding glow voltage, and corresponding glow current respectively. These results were taken in several different experiments, and it will be seen that there is a fairly wide spread. However, if the boundary lines are drawn they show fairly definite trends in direction.

Fig. 8.



Critical discharge voltage.

Fig. 9.



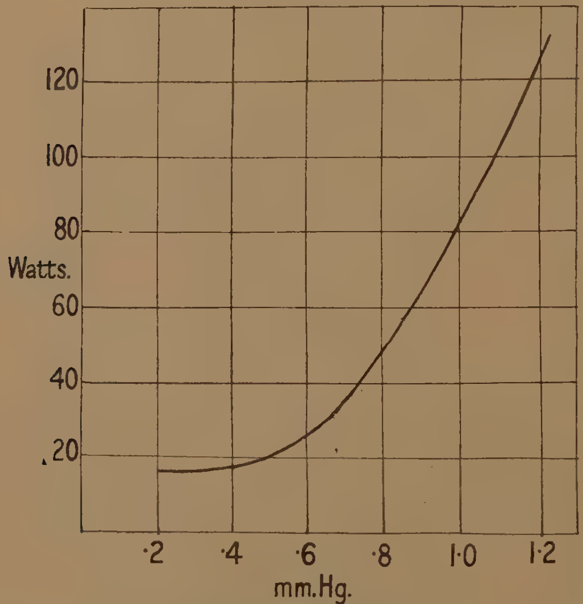
Critical discharge current.

5. *Theories of Transition.*

Possible causes initiating the transition can be divided into three headings, and the data of figs. 8 and 9 examined to see what clue is given by the results.

(a) *Thermal.*—If the breakdown is to be attributed to thermionic emission from the cathode or any other thermal process it is natural to examine the watts consumed in the discharge at the critical points since there is a strong presumption that a large part of these watts pass, to the cathode. These are plotted in fig. 10, and are seen to increase rapidly with the pressure. In order to maintain the criterion of temperature it would therefore be necessary to suppose that the rate of cooling increased after the manner of the figure. The rate of cooling of the electrode will of course increase with gas pressure due to improved convection. We should, however, if that were the explanation, expect the curve to be asymptotic to the axis of watts, as in the Pirani vacuum gauge, rather than to the axis of pressure.

Fig. 10.



Critical discharge watts.

(b) *Field.*—In the case of glow discharge it can be shown * that with certain simplifying assumptions the relationship

$$\frac{IpD^3}{V^2} = \text{constant} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

should hold,

---

* See Wheatcroft, 'Gaseous Electrical Conductors,' chap. vi.

where I = current,

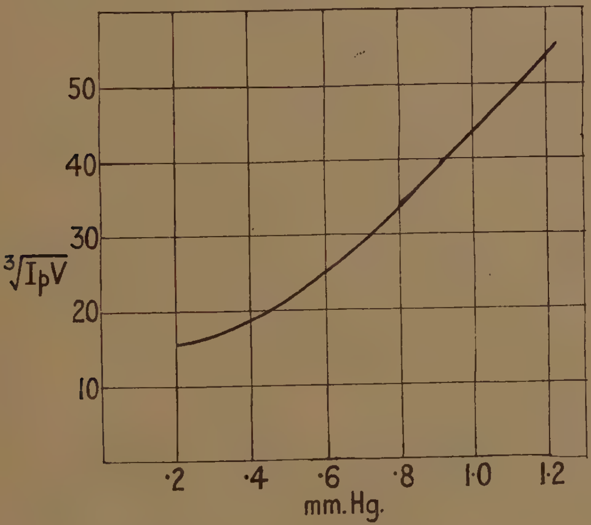
p = pressure,

V = voltage,

D = length of Crookes dark space.

This relation holds provided that the positive ions form the major part of the current at the cathode. The ratio of electron current to ionic current depends upon X/p *, and might be expected to be small if the glow current density is not too large. Equation (1) was established experimentally by Aston † for pressures up to 0·6 mm. and currents of 0·25 milliamps/cm.^{2}. His constant in equation (1) for air was $6{\cdot}4 \times 10^{-11}$ for I in amps/cm.^{2}, p in mm. of mercury, D in cm., and V in volts. Other experiments ‡ by one of us gave a constant of 8×10^{-11}, valid up to 1 mm. pressure and 0·7 milliamps/cm.^{2}.

Fig. 11.



Critical cathode field-strength.

By eliminating D we get X_c proportional to $(IpV)^{\frac{1}{3}}$. $(IpV)^{\frac{1}{3}}$ is plotted in fig. 11. It is seen that the curve still rises with increasing pressure. In order then to maintain the cathode field as the criterion of breakdown, it would be necessary to suppose that $(I/i)^{\frac{1}{3}}$ increases with

---

* Townsend, 'Electricity in Gases.'

† Proc. Roy. Soc. A, lxxix. p. 80 (1907).

‡ Leeds Phil. Soc. ii. p. 483 (1933).

pressure in the same way. But if X is constant X/p should decrease, and so should I/i. Therefore the field theory seems untenable as an explanation of the breakdowns at any rate in this simple form.

(c) *Cathode bombardment.*—The quantity which changes least with increased pressure is the glow voltage as depicted in fig. 8. It may be taken that this voltage is equal to the cathode fall, and so there is a *prima facie* case for the view that the breakdowns are caused by positive ion bombardment, and depend only on the energy of the fastest ions. The suggestion given above will, however, only be tenable if it can be shown that a sufficient number of ions arrive with full energy in the time of voltage rise to breakdown. The length of the dark space may be estimated by the method given elsewhere *. From the data given in that reference, which were also verified during the course of this work, the value of the product of pressure and dark space at the mean voltage of fig. 8 is nearly 0·1 cm. × mm. From this value the lengths of the dark spaces can be calculated; they are inversely proportional to the pressure. The mean free paths of positive ions, if taken to be approximately the same as those of gas molecules, are also inversely proportional to the pressure, so that a constant ratio will be preserved between the dark space lengths and the mean free paths.

The fraction of positive ions reaching the cathode without colliding in travelling across the dark space is given by

$$\frac{N}{N_0}=e^{-D/\lambda},$$

where D = dark space length,

λ = mean free path,

N = number of positive ions reaching cathode,

N_0 = number of positive ions leaving negative glow.

This fraction is a constant, so that the cathode will be bombarded by a number of positive ions possessing full 5 kv. energy, the number being proportional to the current.

The number of ions arriving at the cathode with full energy can then be estimated for these experiments as follows:—

From fig. 9 the current at which transition commences at, say, 0·8 mm. pressure is found to be about 10 mA., which corresponds to a value of $6{\cdot}25\times10^{16}$ ions/sec., if it is composed only of singly-charged ions. When the pressure is 0·8 mm. the mean free path of a positive ion is approximately 10^{-2} cm. From the Paschens-Law curve given in the reference * the value of pD for which the striking and maintenance voltages are 5000 is 0·1 approx., and so the dark space length D is $1{\cdot}25\times10^{-1}$ cm.

---

* Phil. Mag. xx. p. 578 (1935).

Thus

$$N = 6{\cdot}25 \times 10^{16} \times e^{-12\cdot5}$$

$$= 6 \times 10^{10} \text{ ions/sec. (approx.)},$$

showing that some 10^6 ions will arrive at the cathode with full energy every 10^{-4} seconds, which is approximately the most rapid time of voltage rise.

The theory that breakdowns are connected with ion bombardment and probably with cathode disintegration is attractive in view of their temporary nature as seen on the oscillograph, and their localized form as evidenced by scintillations. Conclusive proof can, however, only come from further experiment with this object in view.

II. *A Note on the Schottky Effect.*

By E. L. E. Wheatcroft *.

[Received April 4, 1939.]

The problem of determining I_0, the total emission of a thermionic cathode with zero accelerating field, is of importance in any experiments designed to test emission theory. Since the field is in all cases distorted by negative space charges and there are as yet no quantitative data available for calculating this distortion, the problem resolves itself into that of extrapolating back from measurements made in the region where the distortion is presumably comparatively small, *i. e.*, at high accelerating potentials.

According to Schottky the basis for this extrapolation is the formula

$$\log I = \log I_0 + b\sqrt{V},$$

where b is a constant inversely proportional to the emission temperature. Gill, as a result of tests made on three valves, has challenged the usefulness of the Schottky formula †, and proposed as an alternative the empirical relation

$$I = I_0(1 + b\sqrt{V}).$$

He says that when his experimental points were plotted, the plot of I versus $\sqrt{V}$ "gave a better straight line, but the differences are so small that it would be imperceptible on a small diagram."

I have found in testing two Mazda B 11 triodes that they are particularly suitable for discrimination between the two formulæ. The vacuum is good up to 1800 volts, and at this voltage the field intensity is high because the filament is of small diameter and the grid (which for these tests is the effective anode) is placed close to it. As a result the range of variation of saturation current from say 200 to 1800 volts is large compared with the experimental error.

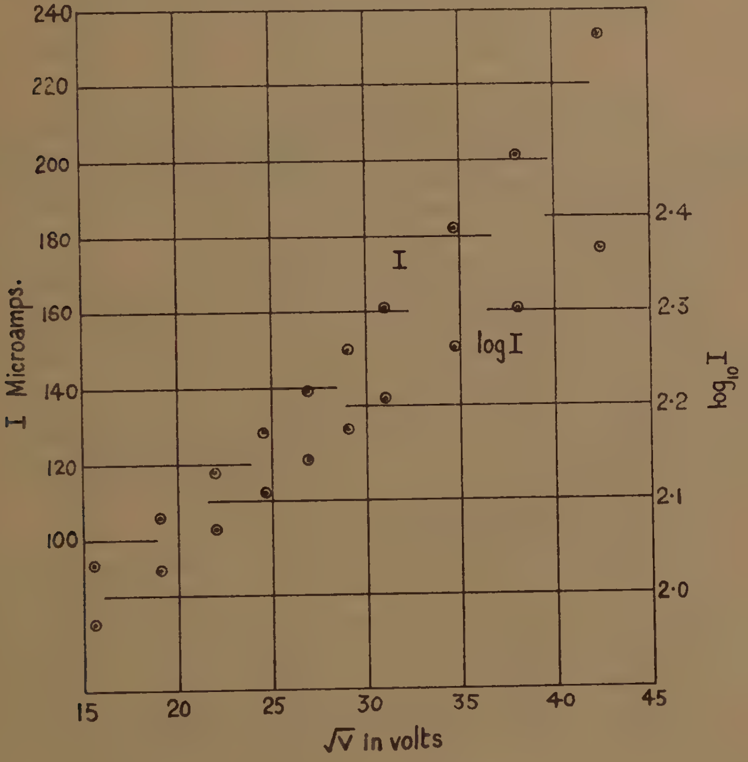
Results of a typical test are shown in the figure. The two valves were each tested at various filament temperatures, and the results were in all cases clearly in favour of Schottky.

Experimental accuracy better than $\pm\frac{1}{2}$ per cent. in the current readings was obtained as follows. The filament was heated by alternating current derived from a Raytheon regulator, which minimized random variations of mains voltage. The filament control was adjusted to a given reading of the microammeter with 1800 volts on the valve. This adjustment was checked between each pair of readings, so that the results were very consistent.

---

* Communicated by the Author.

† Phil. Mag. xxiv. p. 1093 (1937).



Because of space charges the experimental results should be expected to approach the theoretical curve asymptotically from below as the voltage increases, and this presumably accounts for the slight negative curvature of the log plot. I do not see how it is possible to account for the positive curvature found in the plot of I versus $\sqrt{V}$, and it would seem that the onus of proof is therefore still on Gill.

III. *Magnetic Screening by Hollow Circular Cylinders.*

By J. C. Jaeger, Lecturer in Mathematics in the
University of Tasmania *.

[Received January 4, 1939.]

1. Investigations of magnetic screening by cylinders seem to have been confined to steady fields and the steady state effects of sinusoidal fields. In this note the growth of magnetic field inside a hollow conducting cylinder will be studied for some aperiodic external fields. In § 2 the field inside the cylinder is determined for constant external field, in § 3 for an impulsive field, and in §§ 4 and 5 for fields $H_0 e^{-\kappa t}$ and $H_0 \cos \omega t$, all these being for the longitudinal case in which the external field is parallel to the generators of the cylinder. The corresponding problems for the transverse case, in which the external field is perpendicular to the generators of the cylinder, are treated in §§ 6 and 7. Inclined fields may be discussed by superposing the two cases.

Displacement currents are neglected throughout. The method used is that of the Laplace Transformation †.

2. *An infinite, hollow, circular, conducting cylinder $a<r<b$ is free from electric and magnetic fields. At $t=0$ a uniform magnetic field H_0 is established outside the cylinder and parallel to its axis. To find the field in the hollow interior.*

The Maxwell equations, in cylindrical coordinates with z-axis along the axis of the cylinder, give, neglecting displacement currents,

$$-\frac{\partial H_z}{\partial r}=\frac{4\pi\sigma}{c}E_\theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\frac{\partial(rE_\theta)}{r\partial r}=-\frac{\mu}{c}\frac{\partial H_z}{\partial t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where we have used the fact that all quantities are independent of both z and θ. H_z and μ are measured in electromagnetic units, E_θ and the conductivity σ in electrostatic units.

These are to be solved in $a<r<b$ with boundary condition

$$H_z=H_0, \quad \text{when } r=b. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

---

* Communicated by the Author.

† For an exposition *ab initio* of the approach used here see Carslaw, Math. Gazette, xxii. pp. 264–280 (1938).

To find the boundary condition at the inner surface we notice that if displacement currents are neglected H_z will be independent of r for $r<a$. Thus, since H_z and E_θ are continuous at $r=a$, and (2) holds for $r<a$ with $\mu=1$, we get, on multiplying (2) by r and integrating with repect to r from 0 to a,

$$E_\theta=-\frac{a}{2c}\frac{dH_z}{dt} \quad \text{when } r=a, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and this is the required boundary condition.

We now apply the Laplace Transformation to the differential equations (1) and (2) and boundary conditions (3) and (4). Multiply these by e^{-pt}, $p>0$, and integrate with respect to t from 0 to ∞, then, remembering that the initial values of E_θ and H_z are zero, we get the "subsidiary equations"

$$-\frac{d\overline{H}_z}{dr}=\frac{4\pi\sigma}{c}\overline{E}_\theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$-\frac{d(r\overline{E}_\theta)}{rdr}=\frac{\mu p}{c}\overline{H}_z, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

with boundary conditions

$$\overline{H}_z=H_0/p, \quad \text{when } r=b, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$\overline{E}_\theta=-\frac{ap}{2c}\overline{H}_z, \quad \text{when } r=a, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where $\overline{H}_z=\int_0^\infty e^{-pt}H_z\,dt$ is the Laplace Transform of H_z. This "bar" notation will be used throughout.

Eliminating $\overline{E}_\theta$ from (5) and (6) gives

$$\frac{1}{r}\frac{d}{dr}\left[r\frac{d\overline{H}_z}{dr}\right]-q^2\overline{H}_z=0, \quad a<r<b, \quad . \quad . \quad . \quad . \quad (9)$$

where

$$q^2=p/k \quad \text{and} \quad k=c^2/(4\pi\sigma\mu). \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The solution of (9) is

$$\overline{H}_z=AI_0(qr)+BK_0(qr), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and using (5) and substituting in (7) and (8) we find for A and B,

$$\left.\begin{aligned}&AI_0(qb)+BK_0(qb)=H_0/p,\\&A[2\mu I_0'(qa)-aqI_0(qa)]+B[2\mu K_0'(qa)-aqK_0(qa)]=0.\end{aligned}\right\}$$

Solving and substituting in (11) we find

$$\bar{H}_z = \frac{H_0}{p}\,\frac{I_0(qr)[2\mu K_0'(qa) - aqK_0(qa)] - K_0(qr)[2\mu I_0'(qa) - aqI_0(qa)]}{I_0(qb)[2\mu K_0'(qa) - aqK_0(qa)] - K_0(qb)[2\mu I_0'(qa) - aqI_0(qa)]},$$

$$\text{when} \quad a < r < b. \quad \ldots\ldots \quad (12)$$

Thus, using the Inversion Theorem *, we have

$$H_z = \frac{H_0}{2i\pi}$$

$$\times \int_{\gamma - i\infty}^{\gamma + i\infty} e^{\lambda t}\,\frac{I_0(\nu r)[2\mu K_0'(\nu a) - a\nu K_0(\nu a)] - K_0(\nu r)[2\mu I_0'(\nu a) - a\nu I_0(\nu a)]}{I_0(\nu b)[2\mu K_0'(\nu a) - a\nu K_0(\nu a)] - K_0(\nu b)[2\mu I_0'(\nu a) - a\nu I_0(\nu a)]}\,\frac{d\lambda}{\lambda}, \quad \ldots \quad (13)$$

where

$$\nu = \sqrt{(\lambda/k)}. \quad \ldots\ldots\ldots \quad (14)$$

To evaluate (13) consider the integral of its integrand round the closed circuit of fig. 1, which consists of the line AB distant γ from the imaginary axis, and portion of a circle Γ of centre the origin and radius R. R is to be chosen so that Γ does not pass through any pole of the integrand of (13). The integrand is a single-valued function of λ within this contour. It can be shown that the limit of the integral round the circle Γ tends to zero as R tends to infinity through a suitable sequence † of values, avoiding the poles of the integrand of (13). Thus by Cauchy's Theorem the line integral in (13) may be replaced by $2i\pi$ times the sum of the residues of the integrand at its poles within the contour.

There is a simple pole at $\lambda = 0$ of residue 1.

To find the other poles put $\lambda = -k\alpha^2$ in the other factor of the denominator, which becomes

$$\begin{aligned} D(\nu) = D(i\alpha) &= I_0(ib\alpha)[2\mu K_0'(ia\alpha) - ia\alpha K_0(ia\alpha)] \\ &\quad - K_0(ib\alpha)[2\mu I_0'(ia\alpha) - ia\alpha I_0(ia\alpha)] \\ &= -\tfrac{1}{2}i\pi\{J_0(b\alpha)[2\mu Y_1(a\alpha) - a\alpha Y_0(a\alpha)] - Y_0(b\alpha)[2\mu J_1(a\alpha) - a\alpha J_0(a\alpha)]\}. \end{aligned} \quad \ldots \quad (15)$$

---

* That (subject to restrictions on $f(p)$) if

$$f(p) = \int_0^\infty e^{-pt} F(t)\,dt, \quad F(t) = \frac{1}{2i\pi}\int_{\gamma - i\infty}^{\gamma + i\infty} e^{\lambda t} f(\lambda)\,d\lambda,$$

where γ is a real positive constant.

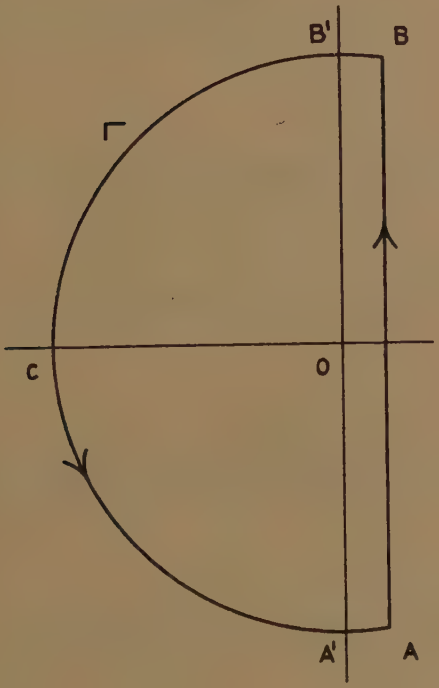
† Such a sequence would be $a\pi(n + \frac{1}{2})/(b - a)$ for large n. The integrand of (13) on these circles then satisfies the conditions of Theorem I. of Carslaw and Jaeger, "A Problem in Conduction of Heat," Proc. Camb. Phil. Soc. (in press). It follows that the integral over Γ tends to zero as $n \to \infty$.

It can be shown that the zeros of (15) are all real and simple. Let these be $\pm\alpha_1$, $\pm\alpha_2$, . . ., then the poles of the integrand of (13) are $\lambda=0$ and $\lambda=-k\alpha_s^2$, $s=1, 2, \ldots$

Using (15) it follows that the residue of the integrand of (13) at $\lambda=-k\alpha_s^2$ is

$$-i\pi e^{-k\alpha_s^2 t}\frac{J_0(r\alpha_s)[2\mu Y_1(a\alpha_s)-a\alpha_s Y_0(a\alpha_s)]-Y_0(r\alpha_s)[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]}{\alpha_s\left[\dfrac{dD(i\alpha)}{d\alpha}\right]_{\alpha=\alpha_s}}, \quad . \; . \; . \quad (16)$$

Fig. 1.



Now from (15), using the recurrence formulæ,

$$\frac{2i}{\pi}\frac{dD(i\alpha)}{d\alpha}=-bJ_1(b\alpha)[2\mu Y_1(a\alpha)-a\alpha Y_0(a\alpha)]+bY_1(b\alpha)[2\mu J_1(a\alpha)-a\alpha J_0(a\alpha)]$$

$$+\frac{1}{\alpha}J_0(b\alpha)\{(2\mu-1)a\alpha Y_0(a\alpha)+(a^2\alpha^2-2\mu)Y_1(a\alpha)\}$$

$$-\frac{1}{\alpha}Y_0(b\alpha)\{(2\mu-1)a\alpha J_0(a\alpha)+(a^2\alpha^2-2\mu)J_1(a\alpha)\}. \quad . \; . \quad (17)$$

Also when $\alpha=\alpha_s$, a zero of (15),

$$\frac{J_0(b\alpha_s)}{2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)}=\frac{Y_0(b\alpha_s)}{2\mu Y_1(a\alpha_s)-a\alpha_s Y_0(a\alpha_s)}=\xi,\ \text{say.} \quad . \quad (18)$$

Introducing this in (17), and using the result

$$J_0(z)Y_1(z)-Y_0(z)J_1(z)=-\frac{2}{\pi z}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

we obtain

$$\frac{2i}{\pi}\left[\frac{dD(i\alpha)}{d\alpha}\right]_{\alpha=\alpha_s}=\frac{2}{\pi\alpha_s}\left\{\xi[4\mu^2-4\mu+a^2\alpha_s^2]-\frac{1}{\xi}\right\}$$

$$=\frac{2}{\pi\alpha_s}\frac{(4\mu^2-4\mu+a^2\alpha_s^2)J_0^2(b\alpha_s)-[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]^2}{J_0(b\alpha_s)[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]}. \quad (20)$$

Thus, finally,

$$H=H_0+\pi H_0$$

$$\times\sum_{s=1}^{\infty}\epsilon^{-k\alpha_s^2 t}\frac{[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]^2\{J_0(r\alpha_s)Y_0(b\alpha_s)-Y_0(r\alpha_s)J_0(b\alpha_s)\}}{(4\mu^2-4\mu+a^2\alpha_s^2)J_0^2(b\alpha_s)-[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]^2},$$

$$\text{when} \quad a<r<b. \quad . \quad . \quad . \quad (21)$$

To find the field in the interior $r<a$ we put $r=a$ in (20), and using (18) and (19), we find,

$$H=H_0-\frac{4\mu H_0}{a}$$

$$\times\sum_{s=1}^{\infty}e^{-k\alpha_s^2 t}\frac{[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]J_0(b\alpha_s)}{\alpha_s\{(4\mu^2-4\mu+a^2\alpha)J_0^2(b\alpha_s)-[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]^2\}},$$

$$\text{when} \quad 0\leqslant r<a \quad . \quad . \quad . \quad (22)$$

When $\mu=1$ this reduces to

$$H_0-\frac{4H_0}{a^2}\sum_{s=1}^{\infty}e^{-k\alpha_s^2 t}\frac{J_2(a\alpha_s)J_0(ba_s)}{\alpha_s^2[J_0^2(b\alpha_s)-J_2^2(a\alpha_s)]}, \quad . \quad . \quad . \quad . \quad (23)$$

where the α_s are the roots of

$$J_0(b\alpha)Y_2(a\alpha)-Y_0(b\alpha)J_2(a\alpha)=0 \quad . \quad . \quad . \quad . \quad (24)$$

From the strict mathematical point of view, since various assumptions have been made in deriving (13), this solution must be regarded as purely formal, and it must be verified that (13) does in fact satisfy the differential equation and boundary conditions of the problem. It is easy to show that the integrand of (13) satisfies the conditions of Theorem 2 of Carslaw and Jaeger, *loc. cit.*, and the verification follows immediately. The derivation of (21) from (13) is rigorous (see previous footnote). The same remarks apply to the other problems considered.

3. *An impulsive field* $H_z = H_0\delta(t)$ *applied outside the cylinder of* § 2.

The Dirac function * $\delta(t)$ is to be regarded as a convenient symbolic representation of a function which is very large in a very small region at the origin and zero elsewhere. It is defined to be zero if $t \neq 0$, and infinite at $t=0$ in such a way † that $\int_0^\infty \delta(t)\,dt = 1.$

It has the property that if $f(t)$ is a continuous function

$$\int_{-\infty}^{\infty} f(t)\delta(t-a)\,dt = f(a),$$

and hence

$$\bar{\delta}(p) = \int_0^\infty e^{-pt}\delta(t)\,dt = 1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Thus, if in the problem of § 2 the external field is $H_z = H_0\delta(t)$ in place of H_0, by (1) above § 2 (7) is to be replaced by

$$\bar{H}_z = H_0, \quad \text{when } r=b.$$

The only change in the solution is that p disappears from the denominator of § 2 (12), and § 2 (13) is replaced by

$$H_z = \frac{H_0}{2i\pi}$$

$$\times \int_{\gamma-i\infty}^{\gamma+i\infty} e^{\lambda t} \frac{I_0(\nu r)[2\mu K_0'(\nu a) - a\nu K_0(\nu a)] - K_0(\nu r)[2\mu I_0'(\nu a) - a\nu I_0(\nu a)]}{I_0(\nu b)[2\mu K_0'(\nu a) - a\nu K_0(\nu a)] - K_0(\nu b)[2\mu I_0'(\nu a) - a\nu I_0(\nu a)]}\,d\lambda, \quad . \quad . \quad . \quad (2)$$

where $$\nu = \sqrt{(\lambda/k)}.$$

The integrand of (2) has no pole at $\lambda=0$; the other poles will be the same as those of § 2 (13), and to find their residues we need simply multiply § 2 (16) by $-k\alpha_s^2$.

Thus we obtain

$$H_z = \frac{4k\mu H_0}{a} \sum_{s=1}^{\infty} e^{-k\alpha_s^2 t} \frac{\alpha_s[2\mu J_1(a\alpha_s) - a\alpha_s J_0(a\alpha_s)]J_0(b\alpha_s)}{(4\mu^2 - 4\mu + a^2\alpha_s^2)J_0^2(b\alpha_s) - [2\mu J_1(a\alpha_s) - a\alpha_s J_0(a\alpha_s)]^2},$$
$$\text{when } 0 \leqslant r < a. \quad . \quad . \quad . \quad (3)$$

---

* *Cf., e. g.*, T. Lewis, Phil. Mag. (7) xxiv. p. 329 (1937).

† It is usual to take $\int_{-\infty}^{\infty} \delta(t)dt = 1$ and $\delta(t)$ to be an even function, but the above is a little more convenient when, as here, we are concerned only with $t \geqslant 0$.

4. *The field* $H_z = H_0 e^{-\kappa t}$ *applied at* $t=0$ *outside the cylinder of* § 2.

Here § 2 (7) is to be replaced by

$$\overline{H}_z = \frac{H_0}{p+\kappa}, \quad \text{when } r=b.$$

Thus the factor λ in the denominator of § 2 (13) is to be replaced by $(\lambda+\kappa)$. There is now a simple * pole at $\lambda=-\kappa$ but none at $\lambda=0$. Then using the results of § 2 we have for $0 \leqslant r < a$,

$$H_z = -\frac{4\mu H_0}{a}$$

$$\times \sum_{s=1}^{\infty} e^{-k\alpha_s^2 t} \frac{\alpha_s J_0(b\alpha_s)[2\mu J_1(a\alpha_s) - a\alpha_s J_0(a\alpha_s)]}{(\alpha_s^2-\beta^2)\{(4\mu^2-4\mu+a^2\alpha_s^2)J_0^2(b\alpha_s)-[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]^2\}}$$

$$-\frac{4\mu H_0}{\pi a\beta} \frac{e^{-k\beta^2 t}}{J_0(b\beta)[2\mu Y_1(a\beta)-a\beta Y_0(a\beta)]-Y_0(b\beta)[2\mu J_1(a\beta)-a\beta J_0(a\beta)]}, \quad . \quad . \quad . \quad (1)$$

where $\kappa = k\beta^2$. . . . . (2)

5. *The field* $H_z = H_0 \cos \omega t$ *applied at* $t=0$ *outside the cylinder of* § 2.

Here § 2 (7) is to be replaced by

$$\overline{H}_z = \frac{pH_0}{p^2+\omega^2}.$$

The factor $\frac{1}{\lambda}$ in the integrand of § 2 (13) is to be replaced by $\frac{\lambda}{\lambda^2+\omega^2}$; the pole at $\lambda=0$ disappears and there are additional poles at $\lambda = \pm i\omega$. Thus in this case, when $0 \leqslant r < a$

$$\overline{H}_z = -\frac{4\mu H_0}{a}$$

$$\times \sum_{s=1}^{\infty} e^{-k\alpha_s^2 t} \frac{k^2\alpha_s^3 J_0(b\alpha_s)[2\mu J_1(a\alpha_s) - a\alpha_s J_0(a\alpha_s)]}{(k^2\alpha_s^4+\omega^2)\{4\mu^2-4\mu+a^2\alpha_s^2)J_0^2(b\alpha_s)-[2\mu J_1(a\alpha_s)-a\alpha_s J_0(a\alpha_s)]^2\}}$$

$$-\frac{2\mu H_0}{\omega' a} \mathbf{R} \left\{ \frac{e^{i\omega t}}{i^{\frac{1}{2}} D(\omega' i^{\frac{1}{2}})} \right\}, \quad , \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $\omega' = \sqrt{(\omega/k)}$ and $D(\nu)$ is defined by § 2 (15).

The first member of (1) is a " transient " term and the formula has been derived to illustrate this. The second member represents the steady-state field *, and can be expressed in terms of the ber and bei functions, but the final formulæ are very clumsy.

---

* It is assumed that κ does not equal any $k\alpha_s^2$.

6. *An infinite hollow † conducting cylinder $a<r<b$ is free from electric and magnetic fields. At $t=0$ a magnetic field, which is uniform at a large distance from the cylinder, is established perpendicular to the axis of the cylinder.*

The z-axis is chosen along the axis of the cylinder, and the x-axis in the direction of the applied field H_0.

Neglecting displacement currents, the field outside the cylinder is given by

$$H_r=-\frac{\partial\phi}{\partial r},\quad H_\theta=-\frac{\partial\phi}{r\partial\theta},$$

where

$$\nabla^2\phi=0. \qquad \ldots \qquad (1)$$

Also H_r and H_θ are to be chosen so that

$$H_r\rightarrow H_0\cos\theta,\quad H_\theta\rightarrow -H_0\sin\theta,\quad \text{as } r\rightarrow\infty. \qquad \ldots \qquad (2)$$

A solution of (1) satisfying (2) is

$$\left.\begin{aligned}\phi&=\left(-H_0r+\frac{A}{r}\right)\cos\theta,\\ H&=\left(H_0+\frac{A}{r^2}\right)\cos\theta,\\ H_\theta&=\left(-H_0+\frac{A}{r^2}\right)\sin\theta,\end{aligned}\right\} \qquad \ldots \qquad (3)$$

where A, a function of t, is to be determined from the boundary conditions at $r=b$.

Similarly in the hollow interior $r<a$ we have a solution,

$$\left.\begin{aligned}H_r&=H\cos\theta,\\ H_\theta&=-H\sin\theta,\end{aligned}\right\} \qquad \ldots \qquad (4)$$

where H is a function of t.

To determine the field within the cylinder $a<r<b$ we use the Maxwell equations in cylindrical coordinates, of which the relevant ones are,

---

* For steady state screening problems see King, Phil. Mag. (7) xv. p. 201 (1933); McLachlan, *ibid.* (7) xviii. p. 610 (1934), also xix. p. 846 (1935).

† The solid cylinder was considered by Lamb, Proc. Lond. Math. Soc. (1) xv. p. 270 (1884).

neglecting displacement currents and remembering that all quantities are independent of z,

$$\left.\begin{aligned} \frac{\partial(r\mathrm{H}_\theta)}{r\partial r}-\frac{\partial \mathrm{H}_r}{r\partial\theta}&=\frac{4\pi\sigma}{c}\mathrm{E}_z,\\ \frac{\partial \mathrm{E}_z}{r\partial\theta}&=-\frac{\mu}{c}\frac{\partial \mathrm{H}_r}{\partial t},\\ -\frac{\partial \mathrm{E}_z}{\partial r}&=-\frac{\mu}{c}\frac{\partial \mathrm{H}_\theta}{\partial t}, \end{aligned}\right\} \quad \ldots\ldots \tag{5}$$

where H_r, H_θ, and μ are measured in e.m.u. and E_z and σ in e.s.u.

Multiplying equations (5) by e^{-pt}, $p>0$, and integrating with respect to t from 0 to ∞, we find the subsidiary equations

$$\frac{\partial(r\overline{\mathrm{H}}_\theta)}{r\partial r}-\frac{\partial \overline{\mathrm{H}}_r}{r\partial\theta}=\frac{4\pi\sigma}{c}\overline{\mathrm{E}}_z, \quad \ldots\ldots \tag{6}$$

$$\frac{\partial \overline{\mathrm{E}}_z}{r\partial\theta}=-\frac{\mu p}{c}\overline{\mathrm{H}}_r, \quad \ldots\ldots \tag{7}$$

$$\frac{\partial \overline{\mathrm{E}}_z}{\partial r}=\frac{\mu p}{c}\overline{\mathrm{H}}_\theta, \quad \ldots\ldots \tag{8}$$

where $$\overline{\mathrm{H}}_\theta=\int_0^\infty e^{-pt}\mathrm{H}_\theta\,dt, \text{ etc.}$$

Eliminating $\overline{\mathrm{H}}_r$ and $\overline{\mathrm{H}}_\theta$ from (6), (7), (8), we have

$$\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial \overline{\mathrm{E}}_z}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2\overline{\mathrm{E}}_z}{\partial\theta^2}-q^2\overline{\mathrm{E}}_z=0, \quad \ldots\ldots \tag{9}$$

where $$q^2=\frac{4\pi\sigma\mu}{c^2}p. \quad \ldots\ldots \tag{10}$$

To obtain a solution of (9) with $\mathrm{H}_r \propto \cos\theta$, let $\overline{\mathrm{E}}_z=\mathrm{X}(r)\sin\theta$, then X must satisfy

$$\frac{d^2\mathrm{X}}{dr^2}+\frac{1}{r}\frac{d\mathrm{X}}{dr}-\left(q^2+\frac{1}{r^2}\right)\mathrm{X}=0, \quad a<r<b.$$

The solution of this is $\mathrm{X}=\mathrm{BI}_1(qr)+\mathrm{CK}_1(qr)$.

Thus $$\overline{\mathrm{E}}_z=[\mathrm{BI}_1(qr)+\mathrm{CK}_1(qr)]\sin\theta \quad a<r<b,$$

and from (7) and (8)

$$\overline{\mathrm{H}}_r=-\frac{c}{\mu pr}[\mathrm{BI}_1(qr)+\mathrm{CK}_1(qr)]\cos\theta, \quad a<r<b, \quad \ldots \tag{11}$$

$$\overline{\mathrm{H}}_\theta=\frac{qc}{\mu p}[\mathrm{BI}_1{}'(qr)+\mathrm{CK}_1{}'(qr)]\sin\theta, \quad a<r<b. \quad \ldots \tag{12}$$

Also multiplying (3) and (4) by e^{-pt}, $p>0$, and integrating with respect to t from 0 to ∞, we have

$$\left.\begin{aligned} \bar{H}_r &= \left(\frac{H_0}{p}+\frac{\bar{A}}{r^2}\right)\cos\theta, \\ \bar{H}_\theta &= \left(-\frac{H_0}{p}+\frac{\bar{A}}{r^2}\right)\sin\theta, \end{aligned}\right\} r>b, \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$\left.\begin{aligned} \bar{H}_r &= \bar{H}\cos\theta, \\ \bar{H}_\theta &= -\bar{H}\sin\theta. \end{aligned}\right\} r<a. \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The boundary conditions are continuity of tangential magnetic field and normal magnetic induction at $r=a$ and $r=b$: these give, using (11), (12), (13), (14),

$$\left.\begin{aligned} -\frac{c}{pb}[BI_1(qb)+CK_1(qb)] &= \frac{H_0}{p}+\frac{\bar{A}}{b^2}, \\ \frac{qc}{\mu p}[BI_1'(qb)+CK_1'(qb)] &= -\frac{H_0}{p}+\frac{\bar{A}}{b^2}, \\ -\frac{c}{pa}[BI_1(qa)+CK_1(qa)] &= \bar{H}, \\ \frac{qc}{\mu p}[BI_1'(qa)+CK_1'(qa)] &= -\bar{H}. \end{aligned}\right|$$

To find the field in the interior we require only $\bar{H}$, and solving, we find

$$\bar{H}=\frac{2b\mu H_0}{ap\Delta(q)},$$

where
$$\begin{aligned} \Delta(q) = {} & [aqI_1'(aq)-\mu I_1(aq)][bqK_1'(bq)+\mu K_1(bq)] \\ & -[aqK_1'(aq)-\mu K_1(aq)][bqI_1'(bq)+\mu I_1(bq)] \quad . \quad (15) \end{aligned}$$

Thus H, the field parallel to the x-axis in the hollow interior, is given by

$$H=\frac{b\mu H_0}{\pi i a}\int_{\gamma-i\infty}^{\gamma+i\infty} e^{\lambda t}\frac{d\lambda}{\lambda\Delta(\nu)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where
$$\nu=\sqrt{\left(\frac{4\pi\mu\sigma\lambda}{c^2}\right)}=\sqrt{\left(\frac{\lambda}{k}\right)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

The integrand of (16) is a single-valued function of λ with a simple pole at $\lambda=0$ of residue

$$\frac{2ab}{b^2(\mu+1)^2-a^2(\mu-1)^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

To find the other poles put $\lambda = -k\alpha^2$ so that $\nu = i\alpha$, then

$$\begin{aligned} \mathit{\Delta}(\nu) = -\tfrac{1}{2}\pi\{&[a\alpha \mathrm{J}_1'(a\alpha) - \mu \mathrm{J}_1(a\alpha)][b\alpha \mathrm{Y}_1'(b\alpha) + \mu \mathrm{Y}_1(b\alpha)] \\ &-[a\alpha \mathrm{Y}_1'(a\alpha) - \mu \mathrm{Y}_1(a\alpha)][b\alpha \mathrm{J}_1'(b\alpha) + \mu \mathrm{J}_1(b\alpha)]\}. \quad . \quad . \quad (19) \end{aligned}$$

The poles of the integrand of (16) are thus $\lambda = 0$ and $\lambda = -k\alpha_s^2$, where α_s, $s = 1, 2, \ldots$ are the zeros of (19). These are all real and simple. The residue at the pole $\lambda = -k\alpha_s^2$ is

$$-\frac{e^{-k\alpha_s^2 t}}{k\alpha_s^2\left[\dfrac{d\mathit{\Delta}(\nu)}{d\lambda}\right]_{\lambda=-k\alpha_s^2}}$$

$$= -\frac{2[a\alpha_s \mathrm{J}_1'(a\alpha_s) - \mu \mathrm{J}_1(a\alpha_s)][b\alpha_s \mathrm{J}_1'(b\alpha_s) + \mu \mathrm{J}_1(ba_s)]e^{-k\alpha_s^2 t}}{\begin{aligned}(\mu^2 - 1 + b^2\alpha_s^2)&[a\alpha_s \mathrm{J}_1'(a\alpha_s) - \mu \mathrm{J}_1(a\alpha_s)]^2 \\ &-(\mu^2 - 1 + a^2\alpha_s^2)[b\alpha_s \mathrm{J}_1'(b\alpha_s) + \mu \mathrm{J}_1(b\alpha_s)]^2\end{aligned}}.$$

Then, as in § 2, the line integral (16) may be replaced by $2i\pi$ times the sum of the residues at its poles, and we have finally

$$\mathrm{H} = \frac{4b^2\mu \mathrm{H}_0}{b^2(\mu+1)^2 - a^2(\mu-1)^2}$$

$$-\frac{4b\mu \mathrm{H}_0}{a}\sum_{s=1}^{\infty}\frac{[a\alpha_s \mathrm{J}_1'(a\alpha_s) - \mu \mathrm{J}_1(a\alpha_s)][b\alpha_s \mathrm{J}_1'(b\alpha_s) + \mu \mathrm{J}_1(b\alpha_s)]e^{-k\alpha_s^2 t}}{\begin{aligned}(\mu^2 - 1 + b^2\alpha_s^2)&[a\alpha_s \mathrm{J}_1'(a\alpha_s) - \mu \mathrm{J}_1(a\alpha_s)]^2 \\ &-(\mu^2 - 1 + a^2\alpha_s^2)[b\alpha_s \mathrm{J}_1'(b\alpha_s) + \mu \mathrm{J}_1(b\alpha_s)]^2\end{aligned}}.$$

$$. \quad . \quad . \quad (20)$$

When $\mu = 1$ we have

$$\mathrm{H} = \mathrm{H}_0 + \frac{4\mathrm{H}_0}{a^2}\sum_{s=1}^{\infty}\frac{\mathrm{J}_2(a\alpha_s)\mathrm{J}_0(b\alpha_s)}{\alpha_s^2[\mathrm{J}_2^2(a\alpha_s) - \mathrm{J}_0^2(b\alpha_s)]}e^{-k\alpha_s^2 t}, \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where the α_s are the roots of

$$\mathrm{J}_2(a\alpha_s)\mathrm{Y}_0(b\alpha_s) - \mathrm{Y}_2(a\alpha_s)\mathrm{J}_0(b\alpha_s) = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

The result (21) is the same as § 2 (22) for the longitudinal case.

7. As in §§ 3, 4, 5, for the longitudinal case, we may deduce from the results of § 6 for the transverse case formulæ for H, the field in the hollow interior, due to other forms of external field.

(i.) External impulsive field $\mathrm{H}_0\delta(t)$:

$$\mathrm{H} = \frac{4b\mu k \mathrm{H}_0}{a}\sum_{s=1}^{\infty}$$

$$\times\frac{\alpha_s^2[a\alpha_s \mathrm{J}_1'(a\alpha_s) - \mu \mathrm{J}_1(a\alpha_s)][b\alpha_s \mathrm{J}_1'(b\alpha_s) + \mu \mathrm{J}_1(b\alpha_s)]e^{-k\alpha_s^2 t}}{\begin{aligned}(\mu^2 - 1 + b^2\alpha_s^2)&[a\alpha_s \mathrm{J}_1'(a\alpha_s) - \mu \mathrm{J}_1(a\alpha_s)]^2 \\ &-(\mu^2 - 1 + a^2\alpha_s^2)[b\alpha_s \mathrm{J}_1'(b\alpha_s) + \mu \mathrm{J}_1(b\alpha_s)]^2\end{aligned}}$$

$$. \quad . \quad . \quad (1)$$

(ii.) External field $H_0 e^{-\kappa t}$:

$$H = -\frac{4b\mu H_0}{a} \sum_{s=1}^{\infty} \frac{\alpha_s^2 e^{-k\alpha_s^2 t}}{\alpha_s^2 - \beta^2}$$

$$\times \frac{[a\alpha_s J_1'(a\alpha_s) - \mu J_1(a\alpha_s)]\,[b\alpha_s J_1'(b\alpha_s) + \mu J_1(b\alpha_s)]}{(\mu^2 - 1 + b^2\alpha_s^2)[a\alpha_s J_1'(a\alpha_s) - \mu J_1(a\alpha_s)]^2 - (\mu^2 - 1 + a^2\alpha_s^2)[b\alpha_s J_1'(b\alpha_s) + \mu J_1(b\alpha_s)]^2}$$

$$-\frac{4b\mu H_0}{\pi a} e^{-k\beta^2 t} \{[a\beta J_1'(a\beta) - \mu J_1(a\beta)][b\beta Y_1'(b\beta) + \mu Y_1(b\beta)]$$

$$-[a\beta Y_1'(a\beta) - \mu Y_1(a\beta)][b\beta J_1'(b\beta) + \mu J_1(b\beta)]\}^{-1}, \quad (2)$$

where $\kappa = k\beta^2$ and β is supposed not equal to any α_s.

(iii.) External field $H_0 \cos \omega t$:

$$H = -\frac{4b\mu H_0}{a} \sum_{s=1}^{\infty} \frac{k^2\alpha_s^4 e^{-k\alpha_s^2 t}}{k^2\alpha_s^4 + \omega^2}$$

$$\times \frac{[a\alpha_s J_1'(a\alpha_s) - \mu J_1(a\alpha_s)]\,[b\alpha_s J_1'(b\alpha_s) + \mu J_1(b\alpha_s)]}{(\mu^2 - 1 + b^2\alpha_s^2)[a\alpha_s J_1'(a\alpha_s) - \mu J_1(a\alpha_s)]^2 - (\mu^2 - 1 + a^2\alpha_s^2)[b\alpha_s J_1'(b\alpha_s) + \mu J_1(b\alpha_s)]^2}$$

$$+ \frac{2b\mu H_0}{a} \mathbf{R} \left\{ \frac{e^{i\omega t}}{\Delta[(i\omega/k)^{\frac{1}{2}}]} \right\} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

8. Numerical calculations from these formulæ are not so formidable as might be expected, particularly now that the British Association tables of Bessel Functions are available. For small values of the time or shell thickness approximate formulæ may be obtained by manipulation of the contour integral expressions, § 2 (13) and § 6 (16). The zeros of § 2 (15) and § 6 (19) are not tabulated, but are readily obtained by interpolation from the B.A. tables.

In this section a few illustrations are given calculated for thick cylinders. All the curves refer to the case $b = 2$ cm., $a = 1$ cm. They represent two materials :

(i.) "Copper," $\mu = 1$, resistivity $1{\cdot}59 \times 10^{-6}$ ohm–cm., so that $k = 126$ cm.²/sec. Here it is convenient to take as unit of time $\tau = 1/126 = {\cdot}00794$ sec.

(ii.) "Iron," $\mu = 5000$, resistivity 12×10^{-6} ohm–cm., $k = {\cdot}19$, $\tau = 1/(0{\cdot}19) = 5{\cdot}23$ sec.

Fig. 2 shows the rise of field in the interior due to unit field applied outside, $H/H_{\text{max.}}$ being plotted against t/τ. Curve I. refers to $\mu = 1$ for both the longitudinal and transverse cases, $H_{\text{max.}} = 1$. Curve II. is the longitudinal case for iron, $\mu = 5000$, $H_{\text{max.}} = 1$. Curve III. is the transverse case for iron, $\mu = 5000$, $H_{\text{max.}} = 16/3\mu$, nearly.

Fig. 2.

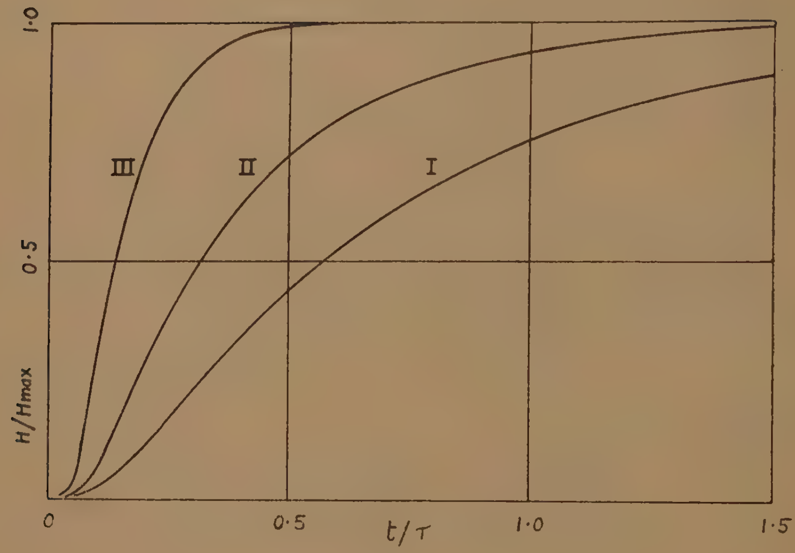
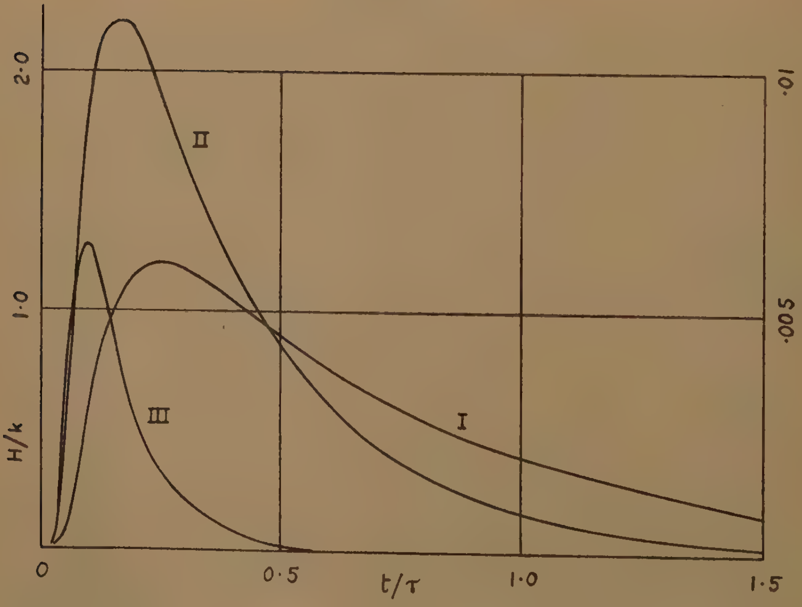


Fig. 3.



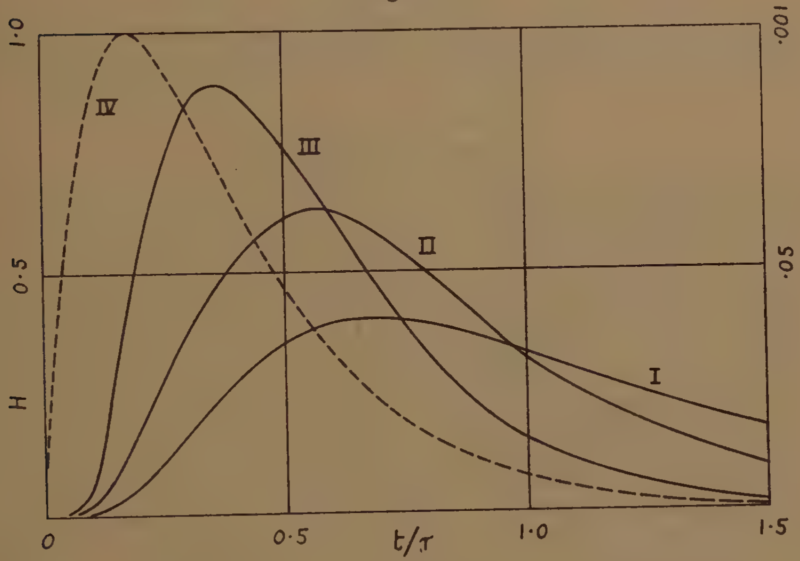
The scale for curves I. and II. is given on the left and that for III. on the right.

In fig. 3 are the internal fields due to unit impulsive field outside. H/k is plotted against t/τ. Curve I. gives both the longitudinal and transverse cases for copper, $\mu=1$. Curves II. and III. are respectively the longitudinal and transverse cases for iron, $\mu=5000$.

In fig. 4 the internal field due to external field,

$$4(e^{-4t/\tau}-e^{-8t/\tau}),$$

Fig. 4.



The scale for curves I., II., and IV. is given on the left and that for III. on the right.

is shown; this field is given in the dotted curve IV. Curve I refers to both longitudinal and transverse cases for copper, $\mu=1$. Curves II. and III. are respectively the longitudinal and transverse cases for iron, $\mu=5000$.

In conclusion, I wish to thank Professor Carslaw for his interest and advice.

IV. *On Iso-dielectric Lines and Geologic Structure.*

By H. Löwy, Dr.phil.*

[Received February 20, 1939.]

The dielectric properties of rocks which are not saturated with water are determined, as I have shown in a previous paper †, by the following equations

$$[\bar{\epsilon}]=v_w+v_c[\epsilon_c], \quad \ldots\ldots\ldots \quad (1)$$

$$v_w+v_a+v_c=1. \quad \ldots\ldots\ldots \quad (2)$$

$\bar{\epsilon}$ and ϵ_c designate the dielectric constant of the rock and the rock-substance, v_w, v_a, and v_c the volume of water, air, and rock-substance contained in 1 c.cm. of the rock. $[\epsilon]$ is an abbreviation for $\frac{\epsilon-1}{\epsilon+2}$.

Formulæ (1) and (2) are the fundamental equations of the *electrodynamic hydrology*.

In the following I discuss some special problems.

I. *Infiltration-depth of Rain.*

If water infiltrates in the soil, the quantities v_w and v_a are changing, whilst v_c remains constant:

$$\Delta v_c=0. \quad \ldots\ldots\ldots \quad (3)$$

From equation (1) it follows that:

$$\Delta[\bar{\epsilon}]=\Delta v_w. \quad \ldots\ldots\ldots \quad (4)$$

Designating by h_r the height of the rain column, by h_w

$$h \;=\Delta v_w, \quad \ldots\ldots\ldots \quad (5)$$

the height of the additional water-column in 1 c.cm. of the rock, we obtain the infiltration depth H

$$\mathrm{H}=\frac{h_r}{h_w}, \quad \ldots\ldots\ldots \quad (6)$$

or, according to (5) and (4),

$$\mathrm{H}=\frac{h_r}{\Delta[\bar{\epsilon}]} \quad \ldots\ldots\ldots \quad (7)$$

On August 16, 1938, measuring ϵ near the well of the rest-house on the desert-road between Cairo and Alexandria, I have found the value

---

* Communicated by the Author.

† H. Löwy, Phil. Mag. xxvii. p. 576 (1939).

$\epsilon=1{\cdot}6$. Several months later, December 9, 1938, returning to the same place, I have measured the same value $\bar{\epsilon}=1{\cdot}6$. For the moment I was astonished because a short rain preceded the measurement. The error of the measurement being less than 0·1, the change of $\bar{\epsilon}$, due to the rain, is at most $\Delta\bar{\epsilon}=0{\cdot}1$. For this value

$$\Delta[\bar{\epsilon}]=[\bar{\epsilon}+\Delta\bar{\epsilon}]-[\bar{\epsilon}]=0{\cdot}023.$$

Therefore, for $h_r=0{\cdot}8$ cm., the minimum infiltration depth is, according to (7),

$$H_{\text{min.}}=35 \text{ cm.}$$

In this calculation we have neglected the evaporation and the superficial discharge of the rain-water. We have supposed the whole quantity of rain infiltrating in the soil.

II. *Water-content of Rocks.*

The quantities in equation (1) are all positive. Therefore

$$v_w \leqslant [\bar{\epsilon}] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

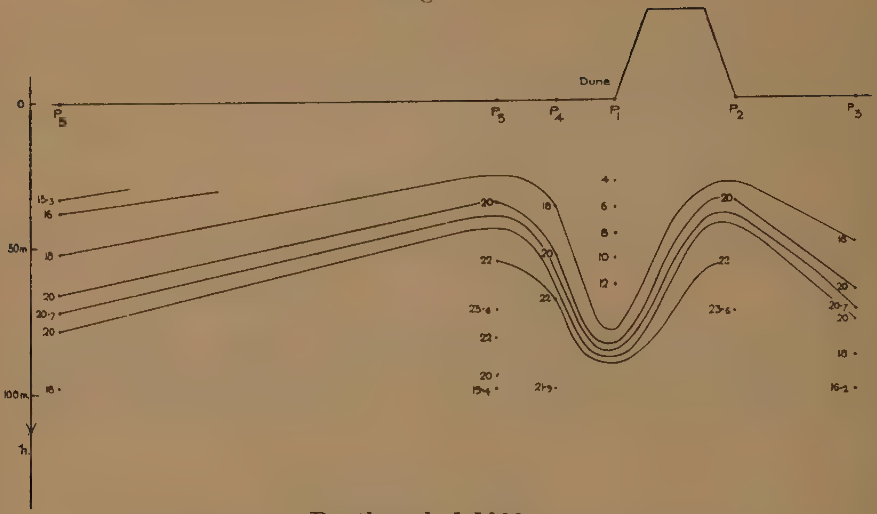
The knowledge of the maximum water-content of rocks, apart from its hydrologic importance, is interesting for the foundation engineer. Before beginning his excavation work he will know whether the water-content of the subsoil has been sufficiently reduced by pumping.

In the following table are contained the values of maximum water-content, which I have determined according to formula (8) near a hill in the Western Desert of Egypt.

Place.	Depth (metres).	Max. v_w (per cent.).	Place.	Depth (metres).	Max. v_w (per cent.).
P_1.........	20	2·3	P_5........	33	19·8
	33	5·1		71	23·6
	71	13·8		98	19·4
P_2.........	33	19·8		136	24·0
	71	23·6	P_6........	33	15·3
P_3.........	33	16·0		71	20·7
	71	20·7		98	18·0
	98	16·2		136	21·5
P_4.........	33	17·6			
	71	22·3			
	98	21·9			

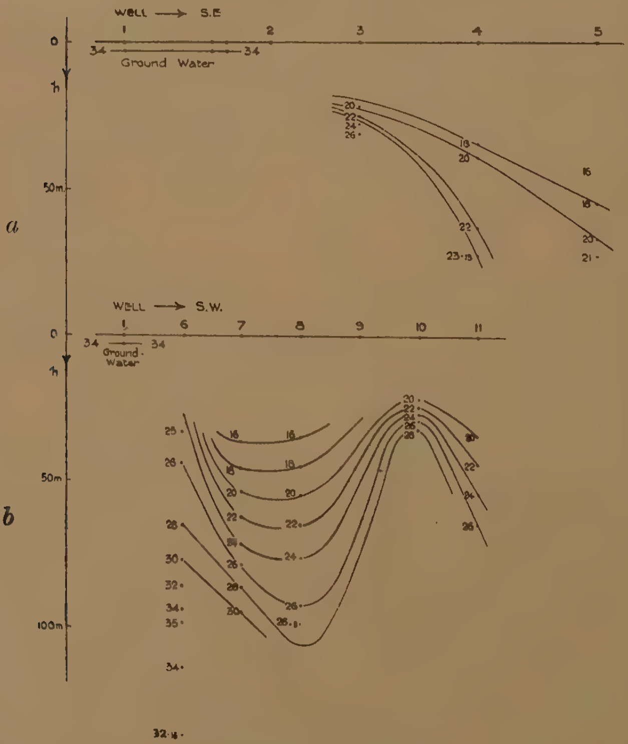
I call "*iso-dielectric*" the lines $[\bar{\epsilon}]=$constant. According to (1), the quantity $\bar{\epsilon}$ is a function of the petrographic (ϵ_c, v_c) and the hydrologic

Fig. 1.


Dune
0
P_5
P_3
P_4
P_1
P_2
P_3
50m
100m
h
15·3
16
18
20
20·7
20
18
20
18
22
20
22
23·4
22
20
15·4
21·9
4
6
8
10
12
20
18
22
23·6
20
20·7
20
18
16·2

Depth-scale 1·1000.
Iso-dielectric lines in the Western Desert of Egypt.

Fig. 2, *a* & *b*.

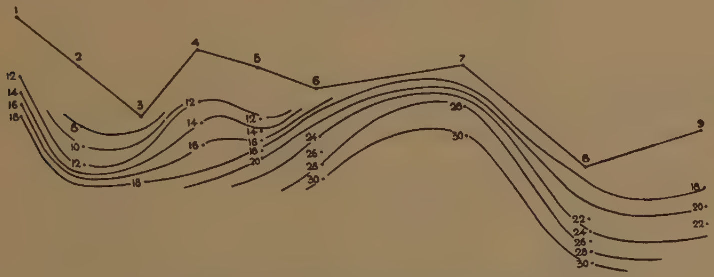

Well → S.E
1
2
3
4
5
0
34
34
Ground Water
h
50m
a
20
22
24
26
18
20
16
18
22
20
23·13
21
Well → S.W.
1
6
7
8
9
10
11
0
34
34
Ground-Water
h
50m
100m
b
25
26
28
30
32
34
35
34
18
18
18
18
20
20
22
22
24
24
26
26
28
30
28·8
20
22
24
26
28
20
22
24
26
32·18

Depth-scale 1·1000.
Iso-dielectric lines in the Western Desert of Egypt.

(v_w) properties of the subsoil. *The iso-dielectric lines, therefore, reveal the hydrologic and petrographic structure.* According to (8), they are *lines of maximum water-content.* In the special case $v_w=0$ they have a purely petrographic character.

In fig. 1 are given, as examples, the iso-dielectric lines, corresponding to the table, in figs. 2 and 3 the lines I have determined in other regions of the Western Desert of Egypt.

Fig. 3.



Depth- and elevation-scale 1·2000.
Iso-dielectric lines in the Western Desert of Egypt.

Fig. 1 shows a syncline under a hill which is situated in a nearly level plane. At the points P_3 and P_6, distant about 3 km., we observe in the depth $h=71$ metres the same sign-change of $\frac{}{\Delta h}$: a line $[\bar{\epsilon}]=20$ is situated above a line $[\bar{\epsilon}]=20{\cdot}7$, and this above another line 20.

In fig. 2, *a* and *b*, the lines reveal a fault or other discontinuity between the well (point 1) and the points 3 and 6, S.E. and S.W. of the well. Between the 4 km. distant points 1 and 6 (fig. 2, *b*) the throw of line $[\epsilon]=34$ amounts to 9 m.

In fig. 3 the iso-dielectric lines follow the surface-profile. The distance between points 1 and 9 is about 50 km.

Hotel Astor,
Cairo (Egypt).
February 12, 1939.

V. *The Electric Field produced by a Point-charge located outside a Dielectric Wedge.*

By S. O. RICE *.

[Received February 1, 1939.]

Introduction.

THE electric field produced by a point-charge lying between two plane conductors which intersect at an angle π/n, where n is any positive integer, may be solved by the method of images. When the angle is not of the form π/n the problem becomes more complicated, and has been investigated by several writers †. Here we shall be concerned with a similar problem. Instead of considering the two intersecting planes to be conductors we assume that they are the two bounding faces of an infinite dielectric of dielectric constant ϵ. The point-charge is outside the dielectric at point Q, as is shown in fig. 1.

The figure shows the cross-section of the system at $z=0$. The cross-section for any other value of z would be the same except for the charge at Q which lies in the plane $z=0$. The line of intersection of the bounding planes is the z-axis. Although the dielectric is drawn in the figure as a wedge, and will be spoken of as such, the results do not require that the angle α be less than β. Two sets of cylindrical coordinates are used, namely, (ρ, θ, z) and (ρ, ϕ, z). The former refer to points outside and the latter to points within the wedge. In accordance with the convention of measuring the angles shown in the figure it follows that $-\beta \leqslant \theta \leqslant \beta$, $-\alpha \leqslant \phi \leqslant \alpha$, and $\alpha+\beta=\pi$.

In all our work we make the assumption that $-\beta<\theta_0<\beta$, that is, the charge does not touch the boundaries of the wedge. Also we assume that the dielectric constant ϵ of the wedge is greater than unity, but is finite. However, if ϵ be set equal to unity or infinity in our series for the potential, the known results will be obtained. This verification is omitted from

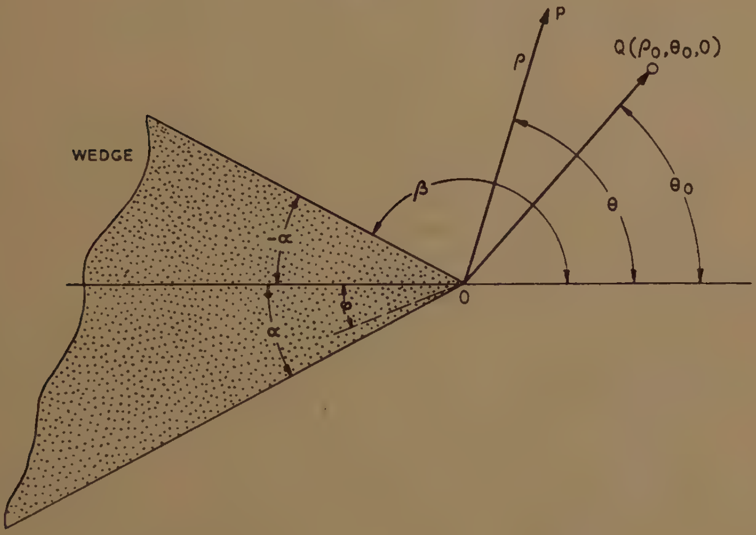
---

* Communicated by the Author.

† F. G. Mehler, *Journal für Reine u. Angewandte Math.* (Crelle's Journal), lxviii. pp. 134–150 (1867) ; H. M. MacDonald, Proc. London Math. Soc. (1) xxvi. pp. 156–172 (1895) ; A. Sommerfeld, Proc. London Math. Soc. (1) xxviii. p. 395 (1897). The first two writers study the associated problem in which the wedge is replaced by a conductor bounded by intersecting spheres. A discussion is also given by H. Bateman, 'Partial Differential Equations of Math. Physics,' Cambridge (1932), pp. 472–473.

this paper since it is relatively straightforward. Our results may be formally extended to the system in which the regions outside and inside the wedge are filled with materials of dielectric constants ϵ_a and ϵ_w respectively. We simply set $\epsilon=\epsilon_w/\epsilon_a$ and multiply the expressions for the potential by $1/\epsilon_a$.

Fig. 1.



Results.

The potential produced by a point-charge q at the point $(\rho_0, \theta_0, 0)$ is

$$V_a=V_0+\frac{iq}{\pi\sqrt{\rho\rho_0}}\,\mathrm{P}\int_{-i\infty}^{+i\infty} Q_{\lambda-\frac{1}{2}}(s)f(\lambda,\theta)\,d\lambda\,. \quad . \quad . \quad . \quad . \qquad (1\,a)$$

outside the wedge (" a " signifying " air ") and

$$V_w=\frac{iq}{\pi\sqrt{\rho\rho_0}}\,\mathrm{P}\int_{-i\infty}^{+i\infty} Q_{\lambda-\frac{1}{2}}(s)g(\lambda,\phi)\,d\lambda \quad . \quad . \quad . \quad . \qquad (1\,w)$$

in the wedge. The path of integration in both integrals is the imaginary axis. Both $f(\lambda, \theta)$ and $g(\lambda, \phi)$ have simple poles at $\lambda=0$, and the letter P standing before the integrals signifies that Cauchy's principal value is to be taken. In these integrals $Q_{\lambda+\frac{1}{2}}(s)$ denotes the Legendre function of the second kind, and the notation used is explained by fig. 1, together with the definitions

$$s=(\rho^2+\rho_0^2+z^2)/2\rho\rho_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

$$V_0=\frac{q}{(\rho^2+\rho_0^2-2\rho\rho_0\cos(\theta-\theta_0)+z^2)^{\frac{1}{2}}}=\frac{q}{\sqrt{2\rho\rho_0}\,[s-\cos(\theta-\theta_0)]^{\frac{1}{2}}}, \qquad (3)$$

$$f(\lambda, \theta)=\frac{(\epsilon-1) \sin 2\lambda\alpha}{2 \sin \lambda\pi}\left[-\frac{\cos \lambda\theta \cos \lambda\theta_0}{D_{\alpha\beta}}+\frac{\sin \lambda\theta \sin \lambda\theta_0}{D_{\beta\alpha}}\right], \quad . \quad (4)$$

$$g(\lambda, \phi)=\frac{\cos \lambda\phi \cos \lambda\theta_0}{D_{\alpha\beta}}+\frac{\sin \lambda\phi \sin \lambda\theta_0}{D_{\beta\alpha}}, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$D_{\alpha\beta}=\epsilon \sin \lambda\alpha \cos \lambda\beta+\cos \lambda\alpha \sin \lambda\beta, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$D_{\beta\alpha}=\epsilon \sin \lambda\beta \cos \lambda\alpha+\cos \lambda\beta \sin \lambda\alpha. \quad . \quad . \quad . \quad . \quad . \quad (7)$$

V_0 is the potential which would be produced by the charge if the wedge were absent.

Since $Q_{\lambda-\frac{1}{2}}(s)$ has a logarithmic singularity at $s=1$, these formulas fail when s has this value, *i. e.*, when we desire the potential at a point on the circle $\rho=\rho_0$, $z=0$. However, it turns out that as the point in question is approached the potential defined by either (1 *a*) or (1 *w*), as the case may be, approaches a limit (unless the point happens to be $[\rho_0, \epsilon_0, 0]$). The potential is defined as this limit.

The expressions (1 *a*) and (1 *w*) for V_a and V_w may be written in a form which does not involve the principal value of an integral by moving the path of integration a distance of $\frac{1}{2}$ to the right. The result is given by equations (16 *a*) and (16 *w*). Closing the paths of integration on the right and evaluating the residues at the poles of $f(\lambda, \theta)$ and $g(\lambda, \phi)$ leads to the series (18 *a*) and (18 *w*) for V_a and V_w.

The corresponding logarithmic potential produced by a line charge of charge q per unit length passing through $(\rho_0, \theta_0, 0)$ parallel to the z-axis is obtained by integrating the point-charge potential along a line of length 2A and then letting A approach infinity. The result is stated in equations (21 *a*) and (21 *w*). The corresponding series expansions, when $\rho<\rho_0$, are given by equations (22 *a*) and (22 *w*).

Conditions to be satisfied by V_a and V_w.

In order that V_a and V_w may represent the potential of a point-charge at $(\rho_0, \theta_0, 0)$ the following conditions must be satisfied :—

1. The potential defined by V_a and V_w and the corresponding normal component of the electric polarization are continuous at the surface of the wedge.

2. V_w satisfies Laplace's equation in the region occupied by the wedge and V_a satisfies the same equation in the region outside the wedge except at $(\rho_0, \theta_0, 0)$. At this point V_a is of the form $V_0+F(\rho, \theta, z)$, where V_0 is the potential due to the point-charge when the wedge is absent, and $F(\rho, \theta, z)$ satisfies Laplace's equation in the region surrounding the point-charge.

3. The quantities rV_a, rV_w are bounded in absolute value for sufficiently large values of r, where r is the distance from the origin, $r^2=\rho^2+z^2$.

Derivation of the Integrals for V_a-V_0 *and* V_w.

The derivation given in this section may be sketched as follows *: First, V_0 is expressed as a suitable infinite integral. It is then assumed that V_a-V_0 and V_w may be expressed as similar integrals with arbitrary multipliers in the integrand. Condition 1 enables us to determine these multipliers, and this completes the derivation.

The integral for V_0 is obtained by expanding its expression (3), using the result †

$$\frac{\pi}{\sqrt{2s-2\cos\omega}}=Q_{-\frac{1}{2}}(s)+2\sum_{n=1}^{\infty}Q_{n-\frac{1}{2}}(s)\cos n\omega. \quad . \quad . \qquad (8)$$

We find in this way that

$$V_0=\frac{q}{\pi\sqrt{\rho\rho_0}}\left[Q_{-\frac{1}{2}}(s)+2\sum_{n=1}^{\infty}Q_{n-\frac{1}{2}}(s)\cos n(\theta-\theta_0)\right], \quad . \quad . \qquad (9)$$

where $s=(\rho^2+\rho_0{}^2+z^2)/2\rho\rho_0$. We shall show that this series is equal to

$$V_0=\frac{q}{\pi\sqrt{\rho\rho_0}}\left\{Q_{-\frac{1}{2}}(s)+i\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}Q_{\lambda-\frac{1}{2}}(s)\frac{\cos\lambda[\pi-|\theta-\theta_0|]}{\sin\lambda\pi}d\lambda\right\}. \qquad (10)$$

The integral converges when $s>1$ and $|\theta-\theta_0|<2\pi$, for when $\theta\neq\theta_0$ the ratio of the trigonometric terms decreases exponentially as $y\to\pm\infty$, $\lambda=\frac{1}{2}+iy$, and when $\theta=\theta_0$ the ratio approaches the constant value i monotonically. At the same time $Q_{\lambda-\frac{1}{2}}(s)$ is of the order of $\xi^{iy}/\sqrt{y}$, ξ being positive, as may be seen from

$$Q_\nu(s)\sim\sqrt{\frac{\pi}{\nu}}\,\xi^{\nu+1}(1-\xi^2)^{-\frac{1}{2}}\{1+0(1/\nu)\}, \quad . \quad . \quad . \quad . \qquad (11)$$

where $\xi=s-\sqrt{s^2-1}$. The asymptotic expression holds when $|\nu|$ is large, $|\arg\nu|<\pi-\epsilon_1$, $|\xi|<1-\epsilon_2$, and $|\arg\xi|<\pi$, where ϵ_1 and ϵ_2 are finite positive quantities however small independent of v ‡. These conditions are all satisfied except at $s=1$.

---

* I wish to thank my colleague, Dr. Marion C. Gray, for pointing out that this method can be applied to the problem. My original derivation, based upon a generalization of a simple two-dimensional case, was less satisfactory because it did not provide a direct method of obtaining the potentials.

† E. W. Hobson, 'The Theory of Spherical and Ellipsoidal Harmonics,' Cambridge (1931) p. 443. The series converges uniformly with respect to s when $s>1$.

‡ E. W. Barnes, "On Generalized Legendre Functions," Quart. Journ. of Math. xxxix. pp. 97–204 (1908). This result is obtained as a special case of §40.

The series (9) is obtained from (10) by closing the path of integration on the right by the three sides of the square having corners at $\left(\frac{1}{2}, \frac{1}{2} \pm i\frac{N}{2}\right)$ and $\left(N+\frac{1}{2}, \frac{1}{2} \pm i\frac{N}{2}\right)$, N being an integer, using the theorem of residues, and then letting $N \rightarrow \infty$. The asymptotic expression (11) applied to $Q_{\lambda-\frac{1}{2}}(s)$ shows that the value of the integral taken along the sides at $y = \pm i\frac{N}{2}$ vanishes at least as fast as $N^{-\frac{1}{2}}$. Furthermore, (11) together with the considerations used in showing the convergence of (10) shows that the integral along $x = N+\frac{1}{2}$ vanishes at least as fast as ξ^{N}. Hence the integral in (10) is proportional to the sum of the residues of the integrand lying to the right of the path of integration, and since the only poles are at the zeros of $\sin \lambda\pi$, it is seen that the series in (9) is obtained.

Since $\sin \lambda\pi$ has no zeros in the strip $0 < R(\lambda) \leqslant \frac{1}{2}$, the path of integration in the expression (10) for V_0 may be shifted to the left by a distance of $\frac{1}{2}$, with the result that the term $Q_{-\frac{1}{2}}(s)$, being π times the residue of the integrand at $\lambda = 0$, combines with the integral to form the principal value of an improper integral,

$$V_0 = \frac{iq}{\pi\sqrt{\rho\rho_0}} \, P\int_{-i\infty}^{+i\infty} Q_{\lambda-\frac{1}{2}}(s) \frac{\cos \lambda[\pi - |\theta - \theta_0|]}{\sin \lambda\pi} \, d\lambda. \quad . \quad . \quad (12)$$

The same argument used in dealing with (10) may be used to show that this integral converges at $\pm i\infty$.

Now that V_0 is expressed in a suitable form we assume that the change in potential due to the presence of the wedge may be expressed as

$$\left.\begin{aligned} V_a - V_0 &= \frac{iq}{\pi\sqrt{\rho\rho_0}} \, P\int_{-i\infty}^{+i\infty} Q_{\lambda-\frac{1}{2}}(s)\,[a(\lambda)\cos\lambda\theta + b(\lambda)\sin\lambda\theta]\,d\lambda, \\ V_w &= \frac{iq}{\pi\sqrt{\rho\rho_0}} \, P\int_{-i\infty}^{+i\infty} Q_{\lambda-\frac{1}{2}}(s)\,[c(\lambda)\cos\lambda\phi + d(\lambda)\sin\lambda\phi]\,d\lambda \end{aligned}\right\} (13)$$

in the air and in the wedge respectively. Since it may be verified by direct differentiation that

$$\rho^{-\frac{1}{2}} Q_{\lambda-\frac{1}{2}}(s)\cos\lambda\theta, \quad \rho^{-\frac{1}{2}} Q_{\lambda-\frac{1}{2}}(s)\sin\lambda\theta$$

are solutions of Laplace's equation, the integrals themselves will be solutions provided that the arbitrary multipliers $a(\lambda)$, $b(\lambda)$, $c(\lambda)$, and $d(\lambda)$ are such (and it turns out that they are) that the order of integration and differentiation may be inverted.

Expressed in equations, condition 1 is

$$\begin{aligned} [V_a]_{\theta=\beta} &= [V_w]_{\phi=-\alpha} \\ \left[\frac{\partial V_a}{\partial\theta}\right]_{\theta=\beta} &= \left[\epsilon\frac{\partial V_w}{\partial\phi}\right]_{\phi=-\alpha} \end{aligned}$$

together with two similar equations obtained by changing the signs of α and β. Applying these conditions to the corresponding integrands in the integral

$$V_a = \frac{iq}{\pi\sqrt{\rho\rho_0}} P \int_{-i\infty}^{+i\infty} Q_{\lambda-\frac{1}{2}}(s) \left[\frac{\cos\lambda[\pi - |\theta-\theta_0|]}{\sin\lambda\pi} + a(\lambda)\cos\lambda\theta + b(\lambda)\sin\lambda\theta \right] d\lambda$$

and in the corresponding expression (13) for V_w leads to four linear equations for $a(\lambda)$, $b(\lambda)$, $c(\lambda)$, and $d(\lambda)$ which may be solved simultaneously. The work is straightforward, and leads directly to the expressions (1 *a*) and (1 *w*) for V_a and V_w.

The convergence of the integrals in (1 *a*) and (1 *w*) may be established with the aid of the inequalities

$$|Q_{\lambda-\frac{1}{2}}(s)| \leqslant \pi(2s-2)^{-R(\lambda)-\frac{1}{2}}, \quad R(\lambda) > -\tfrac{1}{2},\ s>1, \quad . \quad . \qquad (14)$$

$$|f(\lambda,\theta)| < C_1 e^{-|y|(\beta-|\theta_0|)}, \quad |y| > Y, \quad . \quad . \quad . \quad . \qquad (15\,a)$$

$$|g(\lambda,\phi)| < C_2 e^{-|y|(\beta-|\theta_0|)}, \quad |y| > Y, \quad . \quad . \quad . \quad . \qquad (15\,w)$$

where in the last two $y=I(\lambda)$ and Y is a number which depends only upon ϵ and α. C_1 and C_2 are constants which depend only upon ϵ. The first inequality is derivable from

$$Q_{\lambda-\frac{1}{2}}(s) = \int_{-1}^{+1} (1-t^2)^{\lambda-\frac{1}{2}} (2s-2t)^{-\lambda-\frac{1}{2}}\, dt,$$

and the last two may be obtained from the definitions (4) and (5) of $f(\lambda,\theta)$ and $g(\lambda,\phi)$.

V_a *and* V_w *satisfy Laplace's Equation.*

Now that we have completed the derivation, which is based upon condition 1 being satisfied, we seek to verify that the expressions which have been obtained satisfy condition 2. As V_0 is obviously a solution of Laplace's equation, we have only to show that the integrals in (1 *a*) and (1 *w*) are solutions of the same equation. It is not difficult to show that the integrals satisfy Laplace's equation in their respective regions when $s>1$, for then the inequalities (15 *a*) and (15 *w*) for $f(\lambda,\theta)$ and $g(\lambda,\phi)$ and the inequality (14) for $Q_{\lambda-\frac{1}{2}}(s)$ together with similar inequalities for their derivatives show that the differentiations indicated by Laplace's equation may be taken inside the integral sign.

The same kind of considerations shows that the integrals together with all of their first and second partial derivatives with respect to ρ, θ, and z are continuous for all points for which $s>1$ (*i. e.*, for those points which do not lie on the circle $\rho=\rho_0$, $z=0$). However, we still have to dispose of the doubt introduced by the logarithmic singularity of $Q_{\lambda-\frac{1}{2}}(s)$ at

$s=1$. We can do this by showing that the integrals together with all their first and second partial derivatives with respect to ρ,, θ, and z are continuous at $s=1$. The method consists of replacing $Q_{\lambda-\frac{1}{2}}(s)$ in (1 *a*) and (1 *w*) by the first three terms of its expansion around $s=1$ plus a remainder.

$$Q_{\lambda-\frac{1}{2}}(s)=C_0(\lambda, s)+(s-1)(\lambda^2-\tfrac{1}{4})C_1(\lambda, s) \\ +(s-1)^2(\lambda^2-\tfrac{1}{4})(\lambda^2-\tfrac{1}{9})\,C_2(\lambda, s)+I(\lambda, s),$$

where the coefficients are defined by

$$C_n(\lambda, s)=-\frac{1}{2^{n+1}\,n\,!\,n\,!}\left[\psi(\lambda+n+\tfrac{1}{2})+\psi(\lambda-n+\tfrac{1}{2})-2\psi(n+1)+\log\frac{s-1}{2}\right]$$

and, when $R(\lambda)=0$,

$$|\,I(\lambda, s)\,| <(s-1)^{5/2}(A\,|\,\lambda\,|^5+B),$$

$$\frac{\partial}{\partial s}\,I(\lambda, s)<(s-1)^{3/2}(A'\,|\,\lambda\,|^6+B'),$$

$$\frac{\partial^2}{\partial s^2}\,I(\lambda, s)<(s-1)^{1/2}(A''\,|\,\lambda\,|^7+B''),$$

in which A, A′, A″, B, B′, B″ are positive numbers independent of s and λ. Since $f(\lambda, \theta)$ and $g(\lambda, \phi)$ are odd functions of λ, as are also their derivatives (with respect to θ and ϕ), the troublesome logarithmic terms cancel out. The resulting expressions show that the integrals in (1 *a*) and (1 *w*) and all of their first and second partial derivatives with respect to ρ, θ, and z are continuous at $s=1$. The convergence of the integrals involving the logarithmic derivative of the gamma function may be established by (15 *a*) and (15 *w*) (or similar inequalities when the derivatives with respect to θ or ϕ are considered) and the asymptotic expansion *

$$\psi(\xi)\sim\log\xi-\frac{1}{2\xi}+\sum_{n=1}^{\infty}\frac{(-)^n B_n}{2n\xi^{2n}},$$

which holds when $|\,\xi\,|$ is large and $|\arg\xi\,|<\pi$, B_n being the nth Bernoulli number.

The expansion for $Q_{\lambda-\frac{1}{2}}(s)$ stated above may be obtained from Barnes's integral †,

$$Q_{\lambda-\frac{1}{2}}(s)=\frac{1}{4\pi i}\int_C\frac{\Gamma(\lambda+\frac{1}{2}+t)\Gamma(-t)\Gamma(-t)}{\Gamma(\lambda+\frac{1}{2}-t)}\left(\frac{s-1}{2}\right)^t dt,$$

where the path of integration C extends from $-i\infty$ to $+i\infty$, with loops, if necessary, to ensure that the poles of $\Gamma(\lambda+\frac{1}{2}+t)$ lie to the left and those of $\Gamma(-t)$ to the right of the path.

---

* N. Nielsen, 'Handbuch der Theorie der Gammafunktion,' Leipzig, Teubner (1906), p. 208, equation (20).

† E. W. Barnes, *loc. cit.* This result is obtained by setting $m=0$ in the result of § 11 of Barnes's paper.

V_a *and* V_w *at Infinity.*

For the work of this section it is convenient to write the expressions (1 *a*) and (1 *w*) for V_a and V_w in another form, obtained by shifting the path of integration to the right by a distance of $\frac{1}{2}$. It may be shown that neither $D_{\alpha\beta}$ nor $D_{\beta\alpha}$ has zeros in the strip $0<R(\lambda)\leqslant\frac{1}{2}$; hence

$$V_a=V_0-\frac{q(\epsilon-1)\alpha Q_{-\frac{1}{2}}(s)}{\pi(\epsilon\alpha+\beta)\sqrt{\rho\rho_0}}+\frac{iq}{\pi\sqrt{\rho\rho_0}}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}Q_{\lambda-\frac{1}{2}}(s)f(\lambda,\theta)d\lambda, \quad . \quad (16\,a)$$

$$V_w=\frac{qQ_{-\frac{1}{2}}(s)}{(\epsilon\alpha+\beta)\sqrt{\rho\rho_0}}+\frac{iq}{\pi\sqrt{\rho\rho_0}}\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}Q_{\lambda-\frac{1}{2}}(s)g(\lambda,\phi)\,d\lambda. \quad . \quad . \quad . \quad (16\,w)$$

In order to show that condition 3 is satisfied, *i. e.*, that rV_a and rV^i are bounded as r becomes large, r being given by $\sqrt{\rho^2+z^2}$, we note that at large distances from the charge either ρ or z (or both) is large. In any case s must become large because

$$s=\frac{r^2+\rho_0^2}{2\rho\rho_0}\rightarrow\frac{r^2}{2\rho\rho_0}.$$

We examine V_a, as given by (16 *a*), first. The first term, V_0, approaches q/r as $r\rightarrow\infty$. This follows from equation (3) [or the first principles of electrostatic theory]. The second term may be examined by means of the relation

$$Q_{-\frac{1}{2}}(s)=\frac{\pi}{\sqrt{2s}}+0(s^{-3/2}),$$

which is obtained when $Q_{-\frac{1}{2}}(s)$ is expressed in terms of a hypergeometric function of $1/s^2$. The contribution of the last term in (16 *a*) is $0(\rho^{\frac{1}{2}}/r^2)$. This follows from the inequality (14), in which $R(\lambda)=\frac{1}{2}$:

$$|\,Q_{0+iy}(s)\,|<\frac{\pi}{2s-2}\rightarrow\frac{\pi\rho\rho_0}{r^2},$$

and the first of the two inequalities

$$|\,f(\lambda,\theta)\,|<D_1e^{-|y|(\beta-|\theta_0|)} \quad . \quad . \quad . \quad . \quad . \quad (17\,a)$$

$$|\,g(\lambda,\phi)\,|<D_2e^{-|y|(\beta-|\theta_0|)} \quad . \quad . \quad . \quad . \quad . \quad (17\,w)$$

which hold when $R(\lambda)=N+\frac{1}{2}$, N being any integer. In all three of the inequalities $y=I(\lambda)$ may have any value, positive or negative, and in the last two D_1 and D_2 are positive numbers which depend only upon ϵ. The last two may be derived from the definitions of $f(\lambda,\theta)$ and $g(\lambda,\phi)$.

These results show that as $r\rightarrow\infty$

$$V_a\rightarrow\frac{q}{r}-\frac{q(\epsilon-1)\alpha}{\pi(\epsilon\alpha+\beta)}\frac{\pi}{r}=\frac{q}{r}\,\frac{\pi}{\epsilon\alpha+\beta}.$$

When the same reasoning is applied to (16 *w*) it is found that

$$V_w\rightarrow\frac{q}{r}\,\frac{\pi}{(\epsilon\alpha+\beta)}.$$

Thus condition 3 is satisfied. At large distances from the origin V_a and V_w approach the same expression, and the equipotential surfaces become spherical in shape. Moreover, when the total normal electric polarization over a sphere of large radius is computed, using the above approximations for V_a and V_w, it is found to be equal to q, as it should be. This completes the proof that V_a and V_w as given by (16 *a*), (16 *w*), or by (1 *a*), (1 *w*) are the potentials we have been seeking.

Series for the Potentials.

Series for V_a and V_w may be obtained from equations (16 *a*) and (16 *w*) by a method similar to that used to convert the integral (10) for V_0 into the series (9). In each of the integrals the path of integration is closed on the right by the three sides of the square having corners at $\left(\frac{1}{2}, \frac{1}{2} \pm i\frac{N}{2}\right)$ and $\left(N+\frac{1}{2}, \frac{1}{2}, \pm i\frac{N}{2}\right)$, N being an integer. This enables us to apply the inequalities (15 *a*), (15 *w*), (17 *a*), and (17 *w*) for $f(\lambda, \theta)$, and $g(\lambda, \phi)$ at once. These, together with the asymptotic expansion (11) for $Q_{\lambda-\frac{1}{2}}(s)$, show that the contribution of the integral along the three closing sides approaches zero as $N \rightarrow \infty$. Since the poles of $f(\lambda, \theta)$ and $g(\lambda, \phi)$ are all simple, the theorem of residues enables us to write equations (16 *a*) and (16 *w*) as

$$V_a = V_0 - \frac{q(\epsilon-1)\alpha\, Q_{-\frac{1}{2}}(s)}{\pi(\epsilon\alpha+\beta)\sqrt{\rho\rho_0}} + \frac{2q}{\sqrt{\rho\rho_0}} \sum_{n=1}^{\infty} Q_{\lambda_n-\frac{1}{2}}(s) F(\lambda_n, \theta), \quad . \quad (18\,a)$$

$$V_w = \frac{q Q_{-\frac{1}{2}}(s)}{(\epsilon\alpha+\beta)\sqrt{\rho\rho_0}} + \frac{2q}{\sqrt{\rho\rho_0}} \sum_{n=1}^{\infty} Q_{\lambda_n'-\frac{1}{2}}(s) G(\lambda_n', \phi), \quad . \quad . \quad . \quad (18\,w)$$

where $\lambda_1, \lambda_2, \ldots.$ and $\lambda_1', \lambda_2', \ldots.$ are the affixes of the poles of $f(\lambda, \theta)$ and $g(\lambda, \phi)$, respectively, which lie to the right of $R(\lambda)=\frac{1}{2}$. $F(\lambda_n, \theta)$ and $G(\lambda_n', \phi)$ are the residues of $f(\lambda, \theta)$ and $g(\lambda, \phi)$ at the poles $\lambda=\lambda_n$ and $\lambda=\lambda_n'$ respectively. Since it can be shown that the zeros of $D_{\alpha\beta}$ and $D_{\beta\alpha}$ are all real, all these poles lie on the real axis.

Two-dimensional Case.

The potential for the system in which the point-charge at $(\rho_0, \theta_0, 0)$ is replaced by an infinite line-charge parallel to the z-axis and intersecting the plane $z=0$ at $(\rho_0, \theta_0, 0)$ may be obtained from V_a and V_w by integration. The potentials outside and inside the wedge due to a line-charge of finite length extending from $(\rho_0, \theta_0, -A)$ to (ρ_0, θ_0, A) having a uniform charge q per unit length are

$$\int_{-A}^{A} V_a(\rho, \theta, z-z_0)\, dz_0, \quad \int_{-A}^{A} V_w(\rho, \phi, z-z_0)\, dz_0$$

respectively. We wish to obtain the limiting forms of these expressions as A→8.

We first assume that $\rho<\rho_0$ and consider the field outside the wedge. The most convenient form of V_a to substitute in the first of the above integrals is (16 *a*). When z in (16 *a*) is replaced by $z-z_0$ and the resulting expression is integrated with respect to z_0 between $-A$ and $+A$ the first two terms yield, when A is large,

$$\frac{2q\pi}{\epsilon\alpha+\beta}\log 2A+\Phi_0+\frac{2(\epsilon-1)\alpha q}{\epsilon\alpha+\beta}\log\rho_0+0\left\{\frac{(\rho+\rho_0)^2+z^2}{A^2}\right\}, \quad . \qquad (19)$$

where Φ_0 is the logarithmic potential of the line-charge in the absence of the wedge :

$$\Phi_0=-q\log\left[\rho^2+\rho_0^2-2\rho\rho_0\cos(\theta-\theta_0)\right].$$

The integration of the second term is simplified by the use of the expression

$$Q_{-\frac{1}{2}}(s)=\frac{1}{\sqrt{2}}\int_0^\pi\frac{du}{\sqrt{s-\cos u}}.$$

The last term in (16 *a*) gives rise to a double integral, the variables of integration being z_0 and λ. From the inequalities obtained by setting $R(\lambda)=\frac{1}{2}$ in the inequality (14) for $Q_{\lambda-\frac{1}{2}}(s)$ and $N=0$ in the inequality (17 *a*) for $f(\lambda,\theta)$ it may be shown that the double integral approaches a limit as $A\to\infty$. The order of integration may be inverted, by de la Vallee Poussin's test *, in the double infinite integral, and the integral in z_0 evaluated by means of

$$\int_{-\infty}^{+\infty}Q_{\lambda-\frac{1}{2}}\left\{\frac{\rho+\rho_0^2+(z-z_0)^2}{2\rho\rho_0}\right\}dz_0=\frac{\pi\sqrt{\rho\rho_0}}{\lambda}\left(\frac{\rho}{\rho_0}\right)^\lambda,$$

which holds when $R(\lambda)>0$ and $\rho<\rho_0$. This integral may be evaluated by expressing $Q_{\lambda-\frac{1}{2}}(s')$ in terms of a hypergeometric series in $1/s'^2$ and integrating term-wise. Hence the last term in (16 *a*) yields

$$iq\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}\left(\frac{\rho}{\rho_0}\right)^\lambda f(\lambda,\theta)\frac{d\lambda}{\lambda}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (20)$$

The logarithmic potential Φ_a produced outside the wedge by the line-charge may now be obtained by letting A→ 8 in (19) and adding the result to (20). The term in (19) involving log 2A is merely an additive constant. When it is omitted we obtain

$$\Phi_a=\Phi_0+\frac{2(\epsilon-1)\alpha q}{(\epsilon\alpha+\beta)}\log\rho_0+iq\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}\left(\frac{\rho}{\rho_0}\right)^\lambda\frac{f(\lambda,\theta)}{\lambda}d\lambda. \quad . \qquad (21\,a)$$

---

* T. J. I'a Bromwich, 'Theory of Infinite Series,' § 177.

Similarly from equation (16 *w*) it may be shown that the logarithmic potential Φ_w inside the wedge is

$$\Phi_w = -\frac{2\pi q}{\epsilon\alpha+\beta}\log\rho_0 + iq\int_{\frac{1}{2}-i\infty}^{\frac{1}{2}+i\infty}\left(\frac{\rho}{\rho_0}\right)^{\lambda}\frac{g(\lambda,\phi)}{\lambda}\,d\lambda. \quad . \quad . \quad . \qquad (21\,w)$$

Equations (21 *a*) and (21 *w*) are the desired expressions for the logarithmic potentials. Although they have been derived only for $\rho<\rho_0$, they also hold, by the principle of continuation, for $\rho \geqslant \rho_0$, since the integrals converge uniformly for all values of the ratio ρ/ρ_0.

Series for the logarithmic potential analogous to (18 *a*) and (18 *w*) may be obtained by closing the paths of integration of the integrals in equations (21 *a*) and (21 *w*). Thus, if $\rho<\rho_0$, we close the paths on the right and obtain

$$\Phi_a = \Phi_0 + \frac{2q(\epsilon-1)\alpha}{(\epsilon\alpha+\beta)}\log\rho_0 + 2\pi q\sum_{n=1}^{\infty}\left(\frac{\rho}{\rho_0}\right)^{\lambda_n}\frac{F(\lambda_n,\theta)}{\lambda_n}, \quad . \quad . \qquad (22\,a)$$

$$\Phi_w = -\frac{2\pi q}{\epsilon\alpha+\beta}\log\rho_0 + 2\pi q\sum_{n=1}^{\infty}\left(\frac{\rho}{\rho_0}\right)^{\lambda'_n}\frac{G(\lambda'_n,\phi)}{\lambda'_n}. \quad . \quad . \quad . \qquad (22\,w)$$

where the notation is the same as in (18 *a*) and (18 *w*). If $\rho>\rho_0$ we may obtain similar series by closing the paths of integration on the left. These equations are similar in appearance to the expansion of a Green's function in a series of eigenfunctions *.

---

* Courant u. Hilbert, 'Mathematische Physik.'—I., second edition, p. 320.

VI. *A Mechanical Analogy to the Motion of Electrons in Gases.*

By G. D. Yarnold, University College, Nottingham *.

[Received April 21, 1939.]

The steady motion of a stream of electrons through a gas in which a uniform electric field is maintained may be adequately represented for demonstration purposes by the two-dimensioned motion of a steel sphere on an inclined plane into which has been driven at random a large number of nails. The sphere, representing an electron, moves in an irregular manner down the plane, colliding as it does so with the nails which, being fixed, represent the comparatively massive molecules of the gas. Between collisions the sphere is accelerated uniformly, and each free path is seen to be approximately parabolic. Although the model quite clearly exhibits the principal features of the motion of electrons, including lateral diffusion, it is necessary to draw attention to certain points of difference between the actual motion of electrons and the mechanical model.

The fact that in the model the gas molecules are represented as being at rest is of little importance, since in practice the velocities of the electrons are generally many times greater than those of the molecules. There is, however, an important difference between the collisions in the two cases. In the theory of the motion of slow electrons it is general to assume that, immediately following a collision, all directions of motion of an electron are equally probable. This result is a direct consequence of the three-dimensioned nature of the impact and the assumption that the majority of the collisions are perfectly elastic. In the impact between a rolling sphere and a nail there is a greater probability of the sphere being thrown backwards than forwards. Moreover a large part of the energy of the sphere is lost at each collision, and as a result the drift velocity W in the direction of the field is always a considerable fraction of the velocity of agitation u of the moving particle. Nevertheless a statistical study of the motion has yielded results of interest, and the experiments will now be described.

The model which has been used consists of a thick sheet of laminated wood 100 cm. square into which were driven four hundred good quality wire nails each approximately 2 mm. in diameter. To obtain a random distribution of nails the plane was marked out into large squares of side

---

* Communicated by Prof. L. F. Bates, Ph.D., D.Sc.

20 cm. and small squares of side 1 cm. Sixteen nails were allotted to each large square, and the small squares into which they were to be driven were determined by cutting two packs of cards for the coordinates, a hole then being made to mark the position of the nail. If two points came closer together than the diameter of the sphere the second was rejected. The nails were not actually driven into the holes so marked until after the surface had been prepared by rubbing with sand-paper. The sphere used was a ball-bearing approximately 2 cm. in diameter.

Experiments were made with three different inclinations of the plane ($\sin^{-1}$ 0·05, 0·10, and 0·20). The sphere was released in turn ten times from each of twenty equally spaced points in the central section of length 20 cm. of the top edge of the plane, so as to simulate a stream of electrons entering a gas through a slit as in Townsend's well-known experiments(1). After traversing the plane the sphere rolled on to a highly polished horizontal surface, where its velocity of agitation at the instant of leaving the plane was determined approximately. On the average the sphere made about fifteen collisions. In each case the time taken by the sphere to traverse the plane was noted, thus enabling the drift-velocity to be determined. The point at which the sphere left the plane was also recorded, and a measure of the lateral diffusion is given by the ratio R of the number of occasions on which the sphere emerges through the central section of length 20 cm. to the total number of excursions made. The experiments were repeated, using each side of the square in turn as the starting side, and the results given below are therefore based on observations of eight hundred runs for each value of the inclination.

TABLE I.

Inclination.	W (cm./sec.).	u (cm./sec.).	R.
$\sin^{-1}$ 0·05	11·3	30·9	0·323
$\sin^{-1}$ 0·10	15·5	42·4	0·306
$\sin^{-1}$ 0·20	23·0	70·3	0·321

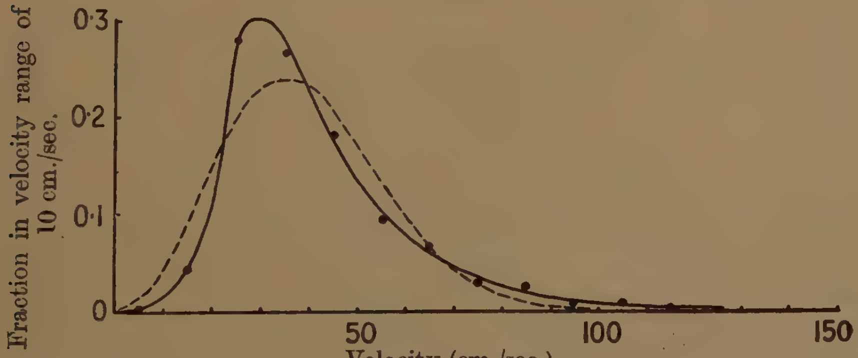
In Table I. are shown the drift velocity W, the root mean square velocity of agitation u, and the ratio R for each inclination. The drift velocity is not proportional to the sine of the angle of inclination, though it does increase with increasing inclination. The theory of the motion of electrons shows, however, that strict proportionality is only to be expected when the drift velocity is small compared with the velocity of agitation. The table also shows that the mean velocity of agitation increases as the inclination increases, while the ratio R appears to be practically constant within the limits of experimental error This is to be expected, since in Townsend's experiments the ratio R increases only very slowly with increase in the electric force.

In Table II. is given a summary of the measurements of the velocity of agitation of the sphere at the two larger inclinations. On account of the effects of friction the velocity measurements at the smallest inclination are not sufficiently reliable to be analysed in this manner. The table shows the number of occasions on which the velocity lay in the ranges 0 to 10 cm./sec., 10 to 20 cm./sec., etc., and in the last column the number of occasions on which the sphere left the plane at the side and its velocity could not be recorded. These same results are also shown by the full curves in figs. 1 and 2, in which the ordinates represent

TABLE II.

Inclination.	Velocity range (cm./sec.).																
	0–10.	10–20.	20–30.	30–40.	40–50.	50–60.	60–70.	70–80.	80–90.	90–100.	100–110.	110–120.	120–130.	130–140.	140–150.	150–160.	Lost.
$\sin^{-1} 0{\cdot}10$	0	31	205	196	132	68	49	21	18	5	5	1	0	0	0	0	69
$\sin^{-1} 0{\cdot}20$	0	0	28	87	146	119	88	69	57	48	33	21	13	8	5	3	75

Fig. 1.

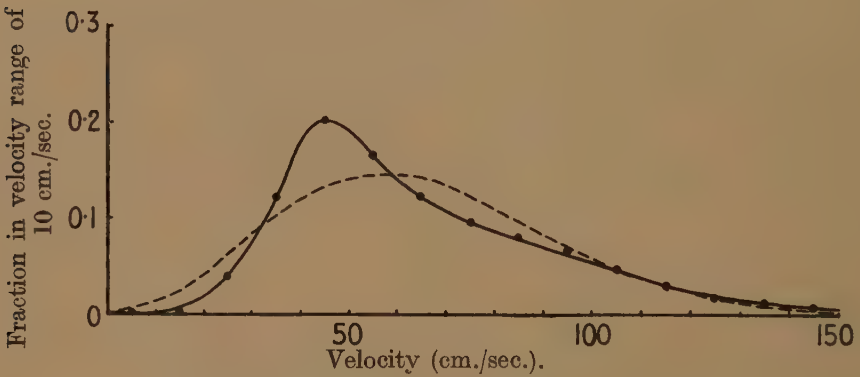


Velocity distribution for inclination $\sin^{-1} 0{\cdot}10$.

the fraction of the total number of occasions on which the velocity lay in the different ranges of 10 cm./sec. Both curves are of the same general shape. In each figure the dotted curve shows the Maxwellian distribution corresponding to the same mean velocity as the experimental curve. The approximate agreement for velocities in excess of the mean is quite remarkable.

The actual velocity distribution of electrons in steady motion in a gas is not yet known for certain, though it has been the subject of a number of investigations. Thus, Langmuir(2) and his collaborators, as a result of observations on glow discharges by the probe method, considered that the velocities are distributed according to Maxwell's Law, while Emeléus, Lunt, and Meek(3), also assuming a Maxwellian distribution together with certain data for single impacts, have calculated values for the coefficient of ionization of electrons in a number of gases which agree quite well with observation. On the other hand, Townsend's theoretical investigations (4). show that the distribution must differ very considerably from the Maxwellian except under conditions of comparatively low electric force and high pressure, in which case the electrons are in approximate thermal equilibrium with the gas molecules.

Fig. 2.



Velocity distribution for inclination $\sin^{-1} 0{\cdot}20$.

It has, however, been shown by Keyston(5) and by Townsend and Jones(6) that the variation with pressure of the intensity of the line spectrum emitted by the positive column of a discharge in helium is in quite good agreement with the assumption of a Maxwellian distribution, but, as was pointed out by Jones(7) in a later paper, this may mean that the velocities of the fastest electrons only, which are responsible for the excitation, are distributed according to Maxwell's Law, while the distribution for the slower electrons are entirely different. In the same way it may be argued that the results of Emeléus, Lunt, and Meek merely establish the Maxwellian distribution for the fastest electrons. Since, however, the closeness of the mechanical analogy described in this paper is open to question it is inadvisable to place much emphasis on the agreement or otherwise with any particular theory of the distribution of velocities.

I should like finally to express my thanks to all those members of the Physics Department who have assisted in making the observations, and in particular to Mr. H. Watson and Mr. R. Blagbrough.

References.

(1) J. S. Townsend, ‘Motion of Electrons in Gases. Clarendon Press (1925).
(2) I. Langmuir and H. M. Mott-Smith, Gen. Elect. Rev. xxvii. p. 538 (1924).
(3) K. G. Emeléus, R. W. Lunt, and C. A. Meek, Proc. Roy. Soc. clvi. p. 394 (1936).
(4) J. S. Townsend, Phil. Mag. ix. p. 1145 (1930); xvi. p. 729 (1933); xxii. p. 145 (1936).
(5) J. E. Keyston, Phil. Mag. xvi. p. 625 (1933).
(6) J. S. Townsend and F. Ll. Jones, Phil. Mag. xv. p. 958 (1933).
(7) F. Ll. Jones, Proc. Phys. Soc. xlviii. p. 513 (1936).

VII. *The Radiation Correction in Calorimetry.*

By F. E. Hoare, Ph.D., A.R.C.S.*

[Received August 4, 1939.]

1. *Introduction.*

There has been proposed a number of different methods of making the so-called radiation correction in calorimetric experiments. Of these the method usually associated with the name of Regnault (1) is probably the most familiar, although it is not necessarily the easiest of application in practice. Rowland (2) proposed an alternative method by means of which the corrections may be made by an entirely graphical method, but in his original paper the underlying assumptions are not set out clearly, and it would appear that there is a mistake of sign which makes his graphical construction, repeated in the 'Dictionary of Physics' (3), erroneous. Of added importance in recent years has been the method of correction to be employed when a complete temperature-time curve is not available, as in the experiments of Eucken (4) at low temperatures and as used for example, by Sucksmith and Potter (5) at high temperatures.

To the best of the author's knowledge there exists no general investigation into the methods of correction and the limits of error permissible in subsidiary determinations for a given limit of error in the corrected temperature rise. It is with the hope of indicating the general conditions of application of these corrections, and the errors likely to arise, that this note is written.

2. *Rowland's Method.*

This is of particular applicability to the method of mixtures. It is assumed that the rate of fall of temperature of the charge dropped into the calorimeter is proportional to the temperature difference existing at any time between the charge and the liquid in the calorimeter. As a corollary the rate of rise of temperature of the liquid will, to a first approximation depending upon the constancy of specific heats, also be proportional to the same difference of temperature.

If the temperature of the charge at any time is represented by T, and

---

* Communicated by Professor F. H. Newman.

that of the liquid by θ, the differential equations representing the rate of change of temperature with time t are, *in the absence of radiation*,

$$\frac{dT}{dt}=-k(\mathrm{T}-\theta),$$

$$\frac{d\theta}{dt}=+\kappa(\mathrm{T}-\theta),$$

where k and κ are constants.

Solving these simultaneous equations, and representing the initial and final temperatures by T_1, T_2 and θ_1, θ_2 respectively, we have

$$\left.\begin{aligned}\mathrm{T}&=\mathrm{T}_2+(\mathrm{T}_1-\mathrm{T}_2)e^{-(k+\kappa)t},\\ \theta&=\theta_2+(\theta_1-\theta_2)e^{-(k+\kappa)t}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

When the calorimeter is losing heat by "radiation" the rate of drop of temperature due to this cause may be written as proportional to the excess temperature of the calorimeter above that of the surroundings provided this is small. Further, as a close approximation if the radiation loss is small, it may be assumed that the actual temperature reached by the calorimeter at the time t is that which it would have reached in the absence of radiation less the integrated effect of the loss of temperature due to radiation. Thus, if the rate of cooling of the calorimeter for unit temperature excess above that of the surroundings is represented by c, and the temperature of the surroundings by T_a, the temperature of the calorimeter at a time t is given by

$$\theta=\theta_2+(\theta_1-\theta_2)e^{-(k+\kappa)t}-\int_0^t c(\theta-\mathrm{T}_a)\,dt. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Since the integral term is small it may be evaluated with sufficient precision by the substitution of the value of θ from (1), *i.e.*, the value of θ in the absence of radiation. We then have

$$\theta=\theta_2+(\theta_1-\theta_2)e^{-(k+\kappa)t}.\left\{\frac{k+\kappa+c}{k+\kappa}\right\}-c(\theta_2-\mathrm{T}_a)t-\frac{c(\theta_1-\theta_2)}{k+\kappa}. \quad . \quad . \quad (3)$$

The maximum value of the temperature, θ_m, occurs at the time t_m given by

$$t_m=\frac{1}{k+\kappa}\log_e\frac{(\theta_2-\theta_1)(k+\kappa+c)}{c(\theta_2-\mathrm{T}_a)},$$

the corresponding value of the temperature being given by

$$\theta_m=\theta_2-\frac{c}{k+\kappa}(\theta_1-\mathrm{T}_a)-c(\theta_2-\mathrm{T}_a)t_m,$$

or

$$\theta_2=\theta_m+c(\theta_2-\mathrm{T}_a)\left\{t_m-\frac{\mathrm{T}_a-\theta_1}{\theta_2-\mathrm{T}_a}\cdot\frac{1}{k+\kappa}\right\}^*, \quad . \quad . \quad . \quad 4)$$

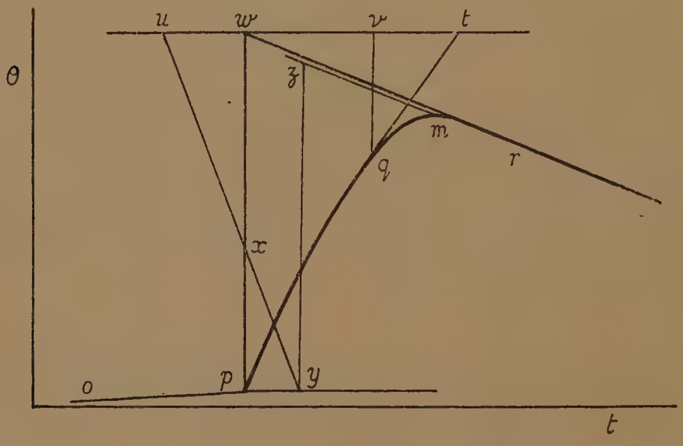
---

* In Rowland's solution the second term in the bracket is prefixed with a positive sign ; that the sign is negative has been carefully examined.

where θ_2 is the temperature which would have been finally attained in the absence of radiation.

The difference between this result and that given by Rowland necessitates a slight alteration in his graphical construction for the corrected temperature rise. Thus in fig. 1 *opqrs* is a plot of the temperature-time curve, and it is required to find the rise of temperature corrected for cooling. Continue the line *sr* until it meets the vertical line *pw* and draw a horizontal line through *w*. At any point such as *q* on the curve draw a tangent and also a vertical line *qv*; the distance *vt* will be very nearly $\frac{1}{k+\kappa}$ of equation (4). Lay off *wu* equal to *vt*, and draw the line *uxy* through the point *x*, which indicates the temperature of the surroundings.

Fig. 1.



To illustrate construction for Rowland's correction.

Draw a vertical line *yz* through *y*. From the point of maximum *m* draw a line *mz* parallel to *sr*, and where it meets *yz* will be the required point, *i. e.*, *yz* gives the rise of temperature corrected for all errors.

The justification for this construction is as follows. At the point *q* the value of θ is given by equation (3) and also

$$\frac{d\theta}{dt}=(\theta_2-\theta_1)e^{-(k+\kappa)t}\{k+\kappa+c\}-c(\theta_2-\mathrm{T}_a).$$

Hence, since
$$vt=qv\bigg/\frac{d\theta}{dt}\doteqdot(\theta_2-\theta)\bigg/\frac{d\theta}{dt}$$
$$=\frac{1}{k+\kappa},$$

neglecting the terms in c.

From this it follows that

$$py=\frac{1}{k+\kappa}\cdot\frac{\mathrm{T}_a-\theta_1}{\theta_2-\mathrm{T}_a}$$

and

$$yz=\theta_m+c(\theta_2-\mathrm{T}_a)\left\{t_m-\frac{\mathrm{T}_a-\theta_1}{\theta_2-\mathrm{T}_a}\cdot\frac{1}{k+\kappa}\right\},$$

since the slope of rs is $-c(\theta_2-\mathrm{T}_a)$ approximately.

3. *Regnault's Method.*

The method due to Regnault for the evaluation of the radiation correction is a numerical or graphical method for the evaluation of the integral of equation (2). This equation may be written more generally as

$$\theta=f(t)-\int_0^t c(\theta-\mathrm{T}_a)\,dt,$$

where $f(t)$ is the temperature which would have been attained at the time t in the absence of radiation.

The purely numerical (6) and combination of graphical and numerical methods (7) for the evaluation of the integral are so well known as to require no detailed description here. The essential process is a step by step evaluation of the integral, using a value of c obtained by observing the rate of change of temperature of the calorimeter before and after the heating process of whatever form. The advantage of this method of correction is that no hypotheses as to the form of the temperature-time curve are necessary, but the numerical work involved is tedious.

4. *Dickinson's Method.*

An ingenious and little-known method of making the correction has been given by Dickinson (7). If the temperature-time curve is of the form shown in fig. 2 a time t_x is found such that the shaded areas of the figure are equal. It is then easy to show that the required correction is given by

$$\int_{t_1}^{t_2} c(\theta-\mathrm{T}_a)\,dt=c(\theta_1-\mathrm{T}_a)(t_x-t_1)+c(\theta_2-\mathrm{T}_a)(t_2-t_x),$$

where t_1 and t_2 are the times corresponding to the temperatures θ_1 and θ_2. The values of c and T_a may be found from the initial and final portions of the temperature-time curve as in Regnault's method.

It is not so difficult to find the required value of t_x as might appear at first sight. For instance, if the temperature-time curve is approximately of the form given by equation (1), as it will be in the method of mixtures, the condition for equality of areas is simply

$$\int_0^{t_x} (\theta_2-\theta_1)(1-e^{-(k+\kappa)t})\,dt=\int_{t_x}^{\infty} (\theta_2-\theta_1)e^{-(k+\kappa)t}$$

or

$$t_x=\frac{1}{k+\kappa}.$$

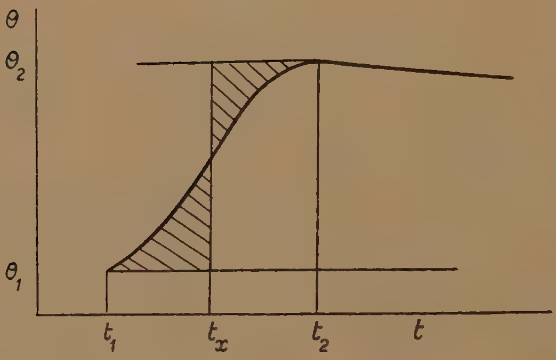
The corresponding temperature is given by

$$\theta_x=\theta_2+(\theta_1-\theta_2)e^{-1}$$

or

$$\theta_x-\theta_1=(\theta_2-\theta_1)(1-e^{-1})$$
$$=0{\cdot}63\ (\theta_2-\theta_1).$$

Fig. 2.



To illustrate Dickinson's method of correction.

Thus the time corresponding to 0·63 of the final temperature rise may be used as t_x. From a large number of experimental determinations with a bomb calorimeter Dickinson found the observed values to lie between 0·57 and 0·62, but states that if for all observations the value 0·60 had been used the error introduced would in no case have exceeded one part in ten thousand.

The case of approximately uniform heating, as, for example, by the dissipation of electrical or mechanical energy at a uniform rate, presents no difficulty in the determination of t_x, for in this case it will lie very nearly at the centre of the heating period.

5. *Method of Nernst and Eucken.*

In the method of calorimetry developed by Nernst and Eucken where a measured amount of electrical energy is supplied to a calorimeter suspended in an evacuated enclosure it is generally not possible to measure the temperature of the calorimeter during the heating process as the same coil is used for the supply of electrical energy as for temperature measurement. When the heating current is switched on the temperature of the heating coil rises rapidly and that of the calorimeter lags very far behind. As a first approximation, however, and for the purposes of calculating the temperature rise corrected for cooling, it may be assumed that the electrical energy is supplied to the calorimeter at a steady rate. In the absence of external heat losses the equation for the rate of rise of temperature may be written

$$\frac{d\theta}{dt}=\mathrm{A},$$

or

$$\theta=\theta_1+\mathrm{A}t, \quad \ldots\ldots\ldots \quad (5)$$

where θ_1 is the initial temperature.

If, as before, it is assumed that the rate of loss of temperature due to "radiation" is proportional to the excess of temperature above that of the surroundings the exact differential equation may be written

$$\frac{d\theta}{dt}=\mathrm{A}-c(\theta-\mathrm{T}_a)$$

or

$$\theta=\theta_1e^{-ct}+\left(\frac{\mathrm{A}}{c}+\mathrm{T}_a\right)(1-e^{-ct}). \quad \ldots\ldots \quad (6)$$

The alternative approximate method of obtaining θ is to write

$$\theta=\theta_1+\mathrm{A}t-\int_0^t c(\theta-\mathrm{T}_a)\,dt.$$

On substituting for θ from (5), and integrating, we have

$$\theta=\theta_1+\mathrm{A}t-c(\theta_1-\mathrm{T}_a)t-\frac{c\mathrm{A}t^2}{2},$$

or, writing $\theta-\theta_1$ for $\mathrm{A}t$ in the last term on the right-hand side, since this is only of the form of a correction term,

$$\theta=\theta_1+\mathrm{A}t+\frac{ct}{2}(\theta_1+\theta-2\mathrm{T}_a). \quad \ldots\ldots \quad (7)$$

It will be seen that equations (6) and (7) are identical to terms of the first order in ct; consequently if (7) is used as the solution for θ it is necessary that ct shall be small in comparison with unity.

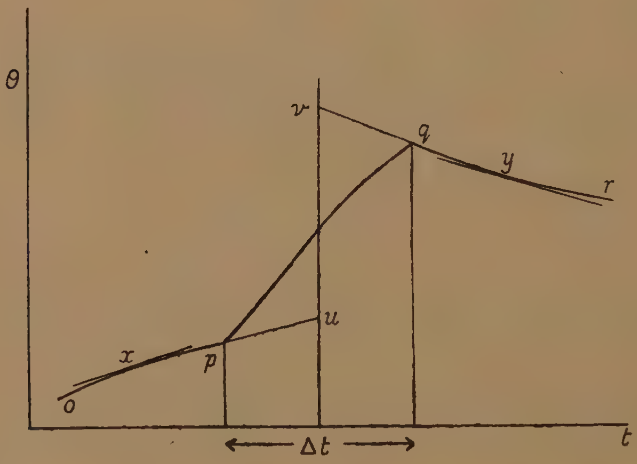
If the heating process is continued for a period Δt the temperature rise, $\mathrm{A}\Delta t$, which would have occurred in the absence of heat losses is

$$\theta_2-\theta_1+\frac{c\Delta t}{2}(\theta_1+\theta_2-2\mathrm{T}_a), \quad \ldots\ldots \quad (8$$

where θ_2 is the actual temperature at the end of the period Δt. When $T_a=\frac{\theta_1+\theta_2}{2}$ the correction to the measured temperature rise vanishes.

Fig. 3 illustrates the results obtained in a typical experiment of the type discussed above. The initial part of the temperature-time curve is represented by *op* and the final part by *qr*. The intermediate part, corresponding to the time during which the heating current flows, is unobtainable if the same coil is used both for heating current and temperature measurements. In this case the experimental measurements of temperature in the immediate neighbourhood of *q* are rather scattered,

Fig. 3.



Typical temperature-time curve for method of Eucken and Nernst.

and some time elapses after the heating current is switched off before the points lie on a smooth curve. The corrected temperature rise is obtained by extrapolating back the line *qr*, provided this is reasonably straight, to the vertical line situated at the middle of the heating period; the initial curve is similarly extrapolated, and if these extrapolated lines cut the vertical in *v* and *u* respectively *vu* is the corrected temperature rise. The advantage of this method, which may be performed either graphically or numerically, is that it does not necessitate a knowledge of the temperature of the surroundings.

In many cases, however, the portions *op* and *qr* of the curve are insufficiently linear for the extrapolations to be made with any degree of confidence. In this case two points *x* and *y* on the initial and final parts of the curve are selected, and by drawing tangents the slopes S_x and S_y at these points are found. Then, since

$$\left.\begin{aligned} S_x&=c(T_a-\theta_x), \\ S_y&=c(T_a-\theta_y), \end{aligned}\right\}, \qquad (9)$$

both c and T_a may be found. The true temperatures θ_1 and θ_2 at the points p and q may then be found, it being assumed that these temperatures have not been obtained experimentally, by the use of an equation of the type

$$\theta-T_a=(\theta_x-T_a)e^{-c(t-t_x)}, \qquad (10)$$

which relates to temperature change due to "radiation" alone. Having found θ_1 and θ_2 equation (8) may be employed for finding the corrected temperature rise.

If the value of $c\Delta t$ is sufficiently small for (8) to be a close enough approximation, a simpler procedure is to extrapolate the initial and final curves to the mid-point of the heating period by the use of equation (10). The difference between these extrapolated temperatures gives the corrected temperature rise to the same order of accuracy as does equation (8), *i. e.*, first power in $c\Delta t$.

A better method giving the corrected temperature rise to a much higher order of accuracy is to solve the exact equation (6) for $A\Delta t$ in ascending powers of $c\Delta t$. Thus

$$A\Delta t=\theta_2-\theta_1+\frac{c\Delta t}{2}(\theta_1+\theta_2-2T_a)+(\theta_2-\theta_1)\frac{c^2\Delta t^2}{12}+\text{terms in } c^4\Delta t^4. \qquad (11)$$

Having obtained θ_2 and θ_1 the use of (11) instead of (8) involves very little extra arithematical work and gives the corrected temperature rise to a much higher order of accuracy.

An alternative method for finding $A\Delta t$ would be to solve equation (6) as it stands for A in terms of θ_1, θ_2, c, and T_a. The temperature rise calculated in this manner should be accurate, but the computation is tedious and, because of the arithmetical manipulations, very liable to error.

The accuracy of the corrected temperature rise depends of course upon the accuracy of the slopes S_x and S_y of equations (9), which are used for the determination of c and T_a. Differentiating (11), after substituting for θ_1 and θ_2 in terms of θ_x, θ_y, etc., we have

$$d(A\Delta t)=-(t_2-t_y)(\theta_2-T_a)\left\{1+\frac{c\Delta t}{2}+\frac{c^2\Delta t^2}{12}\right\}dc+(t_1-t_x)(\theta_1-T_a)$$
$$\times\left\{1-\frac{c\Delta t}{2}+\frac{c^2\Delta t^2}{12}\right\}dc+(\theta_2-T_a)\left\{\frac{\Delta t}{2}+\frac{c\Delta t^2}{6}\right\}dc$$

$$+(\theta_1-\mathrm{T}_a)\left\{\frac{\Delta t}{2}-\frac{c\Delta t^2}{6}\right\}dc-\left\{\frac{\theta_2-\mathrm{T}_a}{\theta_y-\mathrm{T}_a}\left(1+\frac{c\Delta t}{2}+\frac{c^2\Delta t^2}{12}\right)\right.$$

$$\left.-\frac{\theta_1-\mathrm{T}_a}{\theta_x-\mathrm{T}_a}\left(1-\frac{c\Delta t}{2}+\frac{c^2\Delta t^2}{12}\right)\right\}dT_a. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Also from equations (9)

$$\left.\begin{aligned} dc&=\frac{d\mathrm{S}_x-d\mathrm{S}_y}{\theta_y-\theta_x}, \\ d\mathrm{T}_a&=\frac{1}{c}\cdot\frac{\mathrm{S}_x\,d\mathrm{S}_y-\mathrm{S}_y\,d\mathrm{S}_x}{\mathrm{S}_x-\mathrm{S}_y}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Combination of equations (12) and (13) enables dependence of the corrected temperature rise on the slopes S_x and S_y to be determined. This is illustrated below.

6. *Numerical Example.*

As an illustration of the results of the last section the following values have been assumed for a numerical example :

$\theta_1=0°$,
$\mathrm{A}=1°$/min.,
$c=0{\cdot}1°$/min./unit temperature difference,
$\Delta t=3$ minutes,
$\mathrm{T}_a=1°$.

(A large value of c has been chosen to show the order of the approximations more clearly.)

Fig. 4 shows the variation of temperature as calculated from the above values by the use of equation (6).

The maximum temperature reached is 2·8522°.

The correction calculated from (8) is 0·1278°, giving a corrected temperature rise of 2·9800° compared with the exact value of 3·0000°.

The use of equation (11) leads to the value 3·0014°, or an error of 0·047 per cent. compared with that of 0·67 per cent. by the use of (8).

To find the effect of errors in the determinations of the slopes S_x and S_y the conditions three minutes before the heating period commences and three minutes after its close will be considered. We then have

$\theta_x=-0{\cdot}3499$, $\qquad \mathrm{S}_x=0{\cdot}1350$,
$\theta_y=2{\cdot}3719$, $\qquad \mathrm{S}_y=-0{\cdot}1372$.

First let it be supposed that S_x and S_y are each in error by 5 per cent. in such a way as to increase dc.

Then $\qquad dc=0{\cdot}005, \qquad d\mathrm{T}_a=0,$

and $\qquad d(\mathrm{A}\Delta t)=0{\cdot}0278°.$

On the other hand, it may be supposed that one slope is in error in one direction and the other oppositely.

We may then write

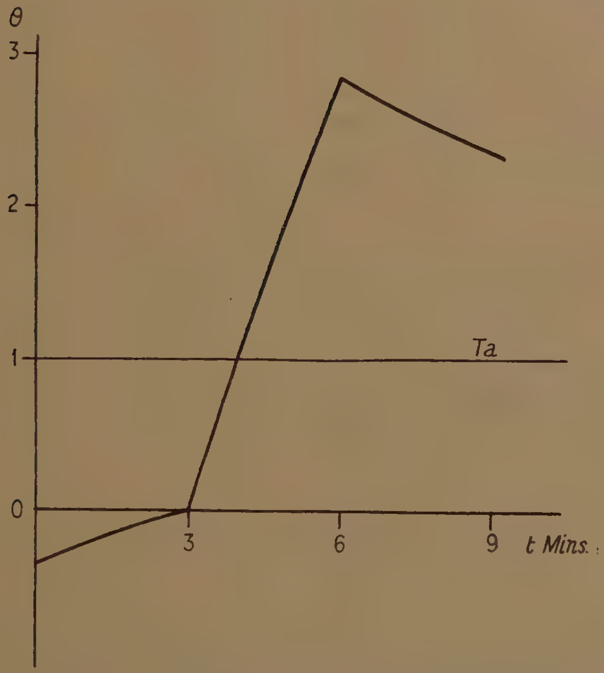
$$dS_x = 0{\cdot}00675, \quad dS_x = 0{\cdot}00686,$$

and on substitution in equations (13) and (12) we have

$$dc = -4{\cdot}04 \times 10^{-5}, \quad dT_a = 0{\cdot}0680,$$

$$d(A\Delta t) = -0{\cdot}0631^\circ.$$

Fig. 4.



Temperature-time curve for typical experiment.

In the above example rather extreme values have been chosen; in an actual experiment it would be arranged that c is not nearly so large compared with A, but the example serves to illustrate that the slopes S_x and S_y should be determined as carefully as possible.

7. *Comparison with Experiment.*

As a test of some of the foregoing results a simple experiment on the cooling of a calorimeter was performed. A calorimeter containing a stirrer and electrical heating coil was mounted centrally in a doublewalled vessel containing water; this served to define the temperature of the surroundings. The temperature of the calorimeter was measured with a mercury-in-glass thermometer calibrated in tenths of a degree Centigrade; hundredths could be estimated easily with a lens.

The following table gives the results obtained in a particular experiment:—

Time (mins.).	Temp. °C.	Time (mins.).	Temp. °C.	Time (mins.).	Temp. °C.
0	11·26	20	14·53	28	19·35
5	11·33	21	15·14	29	19·93
10	11·40	22	15.75	30	19.90
15	11·47	23	16·33	31	19·88
16	12.07	24	16·98	36	19.75
17	12·70	25	17·55	41	19·62
18	13·34	26	18·15	46	19·50
19	13·95	27	18·74		

Current on at 15 mins.; off at 29 mins.
Temperature of surroundings 14·40° C.

These results have been treated by the various methods given above, the paragraph headings giving the references.

§ 3.—Treating the results by the numerical method we have the following:—

$$T_a = 14{\cdot}33^\circ \text{ C.}^*,$$

$$c = 0{\cdot}004726\ ^\circ \text{ C./min./unit temperature difference.}$$

$$\text{Corrected temperature rise} = 8{\cdot}55(3)^\circ \text{ C.}$$

§ 4.—In the case of uniform heating Dickinson's method is equivalent to the use of equation (8). This is given below.

§ 5.—Using the values of c and T_a obtained above from the initial and final parts of the temperature-time curve we have, on substitution in equation (8),

$$\text{Corrected temperature rise} = 8{\cdot}55(1)^\circ \text{ C.}$$

The use of equation (11) gives the result

$$\text{Corrected temperature rise} = 8{\cdot}55(4)^\circ \text{ C.}$$

---

* This is the value obtained by making use of the initial and final portions of the temperature-time curve, and will generally be found to be slightly different from the measured value.

Solving equation (6) for A,

Corrected temperature rise=8·55(2)° C.

This last result is probably less accurate than that obtained from equations (8) or (11) for the reasons already given.

The results were drawn out on squared paper to a large scale, and the initial and final portions of the temperature-time curve produced to the mid point of the heating period. This gave the result

Corrected temperature rise=8·54(5)° C.

There is considerable uncertainty in the last figure, associated with the mere drawing of the lines, so the extrapolations were also made numerically. This led to the result

Corrected temperature rise=8·54(0)° C.

The value given by this method will always be different from the correct result, high or low depending upon the situation of the curve with respect to the temperature of the surroundings. This arises from the linear extrapolation of the cooling curves.

From the above it will be seen that in this particular case there is nothing to choose between the methods of § 3, § 4, equations (8) or (11), or the direct solution of equation (6), although the last is a method not to be generally recommended.

8. *Allowance for Energy Expenditure in Temperature Measurements in Method of Eucken and Nernst.*

Although not coming strictly within the scope indicated by the title of this paper, it is perhaps worth while to use the equations developed above to find the allowance to be made for energy expenditure in realizing the initial and final parts of the temperature-time curve in the method of calorimetry developed by Eucken and Nernst.

If the same coil is used for temperature measurement by resistance thermometry and for carrying the much heavier current during the heating period it is necessary to find to what rate of energy expenditure the temperature rise calculated by the methods given above corresponds.

Let us suppose that the current flowing whilst temperature measurements are being taken would in the absence of heat loss by "radiation" give a rate of rise of temperature of the specimen A_1. Also during the heating period we may suppose the corresponding quantity to be A_2. Then from equation (7) the corrected rise of temperature in the period Δt is given by

$$A_2\Delta t=\theta_2-\theta_1+\frac{c\Delta t}{2}(\theta_1+\theta_2-2T_a). \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Owing to the expenditure of energy in the coil during temperature measurements equations (9) have to be modified, and now become

$$S_x=c(T_a-\theta_x)+A_1,$$

$$S_y=c(T_a-\theta_y)+A_1.$$

Hence $$c=\frac{S_x-S_y}{\theta_y-\theta_x} \text{ as before,}$$

and $$T_a+\frac{A_1}{c}=\frac{S_x\theta_y-S_y\theta_x}{S_x-S_y}=T_a' \text{ say.}$$

If in (14) $T_a'-\frac{A_1}{c}$ is substituted for T_a this becomes

$$A_2\Delta t=\theta_2-\theta_1+\frac{c\Delta t}{2}\left(\theta_1+\theta_2-2T_a'+\frac{2A_1}{c}\right),$$

or $$(A_2-A_1)\Delta t=\theta_2-\theta_1+\frac{c\Delta t}{2}(\theta_1+\theta_2-2T_a').$$

Thus, if equations (9) are solved in the usual way for T_a and c, and the corrected temperature rise calculated from (8) or from (11), it corresponds to that due to increasing the rate of heating from A_1 to A_2, *i.e.*, to the difference of the rates of heating. This result is easily transformed into terms of energy expenditure.

The calculation of the temperatures θ_1 and θ_2 from the observed values of θ_x and θ_y will also be unaffected, for instead of equation (10) we will now have

$$\theta=\theta_x e^{-c(t-t_x)}+\left(\frac{A_1}{c}+T_a\right)(1-e^{-c(t-t_x)}),$$

or $$\theta-T_a'=(\theta_x-T_a')e^{-c(t-t_x)},$$

which is of the same form as equation (10).

Note added in proof.—Another method, due to Ferry, of obtaining the corrected temperature rise is described in the 'Dictionary of Applied Physics' (3). Essentially the method consists of extrapolating the initial and final parts of the temperature–time curve to the vertical line through the temperature T_a on this curve. The corrected temperature rise is then to be taken as the height between these points of interception. There would appear to be little justification for such a procedure in the general case, for the analysis of § 4 shows that to obtain the correction the line to which the extrapolations must be made depends upon the shape of the temperature-time curve, and this depends upon the manner of heating. Applying this method of correction, however, to the two cases considered above, *i. e.*, method of mixtures and uniform heating, it is found that Ferry's method does not give the true correction even as far as the first power in c. If the special case of uniform heating with the temperature T_a equal to the mean of the initial and final temperatures is considered, then Ferry's method obviously becomes equivalent to that discussed in § 5 and the correction vanishes. Only in this case is the method correct to terms of the first power in c.

References.

(1) Regnault, 'A Text-book of Practical Physics,' Watson, p. 218.
(2) Rowland, Proc. Amer. Acad. xv. p. 75 (1880).
(3) 'Dictionary of Applied Physics,' i. p. 62.
(4) Eucken, *Phys. Zeit.* x. p. 586 (1909).
(5) Sucksmith and Potter, Proc. Roy. Soc. cxii. p. 157 (1926).
(6) 'Dictionary of Applied Physics,' i. p. 61.
(7) Dickinson, Sci. Papers Bur. Stds. ii. p. 189 (1914)

VIII. *The Method of Symmetrical Components applied to Harmonic Analysis.*

By Louis A. Pipes, Ph.D.,
The Harvard Engineering School, Cambridge, Mass.*

[Received February 6, 1939.]

Abstract.

A method of determining the Fourier series coefficients from a graphically known wave is presented. The special feature of this method is the use of Fortescue's sequence operator S_m, which plays such a prominent rôle in the theory of Symmetrical Components (1). The determination of the various coefficients is then reduced to a simple matrix multiplication scheme involving the sequence operators S_m and a columnar matrix $[u]$ formed from the ordinates of the graphically known wave.

I. *Introduction.*

In many investigations of applied science it is necessary to carry out a harmonic analysis of graphically known functions. Methods and elaborate schedules for the determination of the Fourier coefficients of such functions are very numerous (2). Hence there appears to be a very small possibility of evolving a new method which would be an improvement over the many existing systems. In this discussion, however, a method is presented which makes use of the same notions that the electrical engineer uses in the theory of Symmetrical Components. By the use of simple matrix algebra the results are expressed in a very convenient form. Since matrix multiplication is so readily effected with a computing machine, the method should prove quite advantageous when such an instrument of calculation is available.

II. *The General Problem.*

The problem in a general form may be stated in the following manner. Given a graphically known function, $u(t)$, of period T, let it be required to find a sum,

$$\sum_{s=-r}^{s=+r} A_s e^{2\pi jst/T}, \qquad \ldots\ldots\ldots \qquad (1)$$

---

* Communicated by the Author.

which furnishes the best possible representation of the function $u(t)$ when we are given that $u(t)$ takes on the values $u_0, u_1, \dots u_{n-1}$ when t takes on the values 0, T/n, $2\mathrm{T}/n$, $\dots (n-1)\mathrm{T}/n$, respectively. If we let $r=2/n$, the problem is to determine the $(2r+1)$ constants $\mathrm{A}_{-r} \dots \mathrm{A}_0 \dots \mathrm{A}_r$ so as to make the expression (1) take, as nearly as possible, the above values of u when t takes on its corresponding values.

III. *Determination of the Coefficients.*

In order to determine the above coefficients A_s, let us introduce the complex number $a=e^{2\pi j/n}$. This complex number may be regarded as a unit vector in the complex plane that makes an angle of $2\pi/n$ with the real axis. Substituting the corresponding values of $u(t)$ and t into the expression (1) we obtain the following equations to determine the A_s:

$$u_0=\sum_{s=-r}^{s=+r} \mathrm{A}_s, \qquad \dots \dots \dots \qquad (2)$$

$$u_1=\sum_{s=-r}^{s=+r} \mathrm{A}_s a^s, \qquad \dots \dots \dots \qquad (3)$$

. . . . .

$$u_p=\sum_{s=-r}^{s=+r} \mathrm{A}_s a^{ps}, \qquad \dots \dots \dots \qquad (4)$$

. . . . .

$$u_{(n-1)}=\sum_{s=-r}^{+r} \mathrm{A}_s a^{(n-1)s}. \qquad \dots \dots \dots \qquad (5)$$

Now to obtain the general coefficient A_m, form the sum

$$[u_0+a^{-m}u_1+a^{-2m}u_2+ \dots +a^{-(n-1)m}].$$

We then obtain

$$\sum_{k=0}^{k=n-1} u_k a^{-km}=\sum_{s=-r}^{s=+r} \mathrm{A}_s[1+a^{(s-m)}+a^{2(s-m)}+ \dots +a^{(n-1)(s-m)}]. \qquad . \quad (6)$$

The expression in brackets at the right of equation (6) represents the sum of n equally-spaced unit vectors provided that $s \neq m$. Such a sum vanishes. When $s=m$, the term in brackets has the value n. It is thus seen that only the term $s=m$ remains, and the right side of equation (6) becomes $n\mathrm{A}_m$. We thus obtain the general result:

$$\mathrm{A}_m=\frac{1}{n}\sum_{k=0}^{k=(n-1)} u_k a^{-km}. \qquad \dots \dots \dots \qquad (7)$$

Formula (7) enables us to find the various coefficients.

IV. *The Sequence Operators* S_m.

To compute the coefficients A_m we make use of the sequence operators S_m used in the Theory of Symmetrical Components(1). The sequence operator, S_m, may be defined as the following line matrix :

$$S_m=[1, a^{-} \ , a^{-2m}, \ldots a^{-(n-1)m}]. \quad \ldots \ldots \quad (8)$$

If we now introduce the column matrix $[u]$, constructed in the following manner,

$$[u]=\frac{1}{n}\begin{Vmatrix} u_0 \\ u_1 \\ u_2 \\ — \\ u_{n-1} \end{Vmatrix} \quad \ldots \ldots \ldots \quad (9)$$

then we have the general coefficient given by the following equation :

$$A_m=S_m[u], \quad \ldots \ldots \ldots \quad (10)$$

in view of formula (7). Now, since the line matrices S_m and S_{-m} are conjugate matrices, we may write

$$S_{-m}=\bar{S}_m, \quad \ldots \ldots \ldots \quad (11)$$

and it follows that

$$A_{-m}=\bar{A} \quad \ldots \ldots \ldots \quad (12)$$

It is thus only necessary to determine the coefficients A_0, A_1, A_2, . . . A_r since the coefficients A_{-1}, A_{-2}, . . A_{-r} are the conjugates of the respective coefficients with positive subscripts.

If we introduce the angular frequency $w=2\pi/T$, we may write

$$u(t)=\sum_{s=-r}^{s=+r} A_s e^{jswt}. \quad \ldots \ldots \ldots \quad (13)$$

In view of (10), this may be written in the form

$$u(t)=\sum_{s=-r}^{s=+r} S_s[u]e^{jswt}. \quad \ldots \ldots \ldots \quad (14)$$

In most physical problems the use of the complex Fourier series has certain advantages over the real Fourier series in terms of sines and cosines.

For example, if we have the function expressed as a series of the form (13), then the effective value of u is given by the following formula :

$$u_{\text{eff.}}=\sqrt{\sum_{s=-s}^{s=+r} A_s A_{-s}}. \quad \ldots \ldots \ldots \quad (15)$$

The complex form of the Fourier series is of particular importance in electrical circuit theory. Many interesting applications of the complex

Fourier series are given in reference (3). If it is desired to express the function in the real Fourier series form,

$$u(t)=a_0+\sum_{s=1}^{r} a_s \cos(swt)+\sum_{s=1}^{r} b_s \sin(swt), \quad . \quad . \quad . \quad (16)$$

where $w=2\,/\mathrm{T}$, then the coefficients a_s and b_s may be obtained from the following equations:

$$a_0=\mathrm{A}_0=\mathrm{S}_0[u], \quad . \quad . \quad . \quad (17)$$

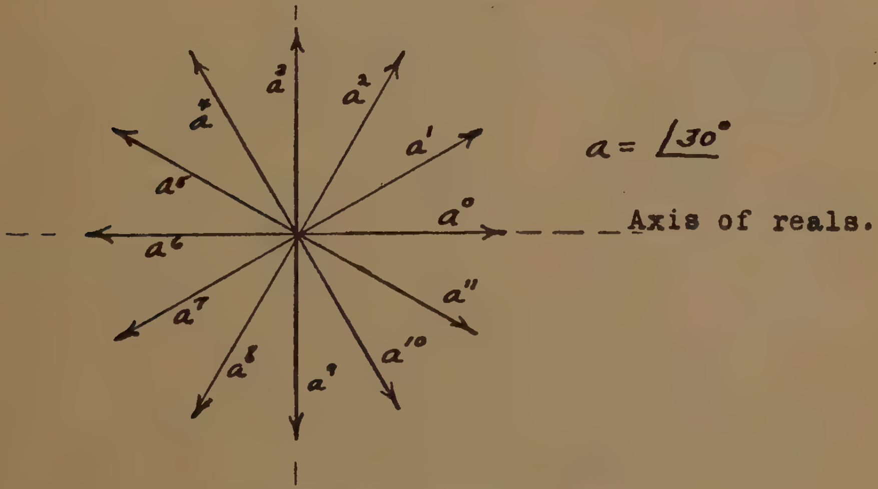
$$a_s=(\mathrm{A}_s+\mathrm{A}_{-s})=(\mathrm{S}_s+\bar{\mathrm{S}}_s)[u] \qquad s\neq 0, \quad . \quad . \quad . \quad (18)$$

$$b_s=j(\mathrm{A}_s-\mathrm{A}_{-s})=j(\mathrm{S}_s-\bar{\mathrm{S}}_s)[u] \qquad s\neq 0. \quad . \quad . \quad . \quad (19)$$

V. *Illustrative Example.*

In most practical cases the number of given values of the graphically known function $u(t)$ is either 12 or 24. Let us consider the case where $n=2r=12$. In this case we know the values u_0, u_1, u_2, . . . u_{11} at $t=0$, T/12, 2T/12, . . . 11T/12, where T is the fundamental period of the function $u(t)$. We then construct the matrix $[u]$ by equation (9). It is then necessary to compute the sequence operators S_0, S_{-1}, . . . S_{-6}. This may be most simply done by means of a polar plot of the vectors a, a^2, a^3, . . . a^{11}, where in this case a is the vector $e^{2\pi j/12}$, that is, in the 12-ordinate scheme a is a unit vector making 30° with the real axis. Such a polar plot is most easily constructed on polar graph paper. From this polar plot the sequence operators may be computed and tabulated. It is evident that once these sequence operators are obtained and tabulated they may be used for any 12-ordinate scheme of analysis, since this part of the work is independent of the function to be analysed. By means

Fig. 1.



Polar plot (12-ordinate method).

TABLE I.

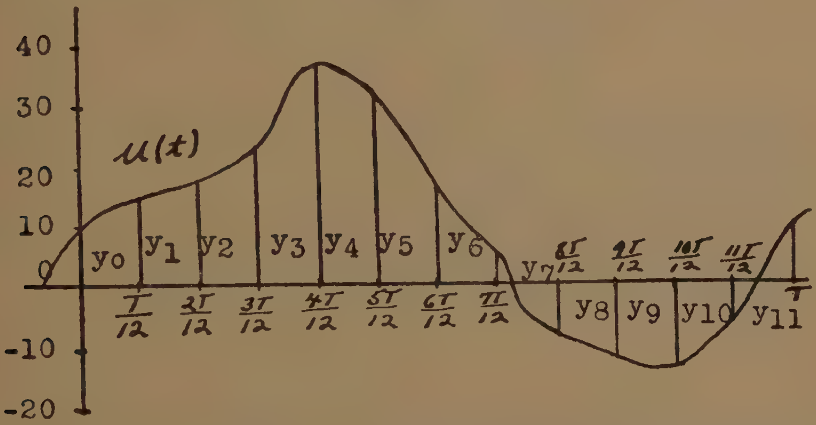
Sequence Operators for 12-ordinate Scheme.

$S_0=$	1	1	1	1	1	1	1	1	1	1	1	1
$S_{-1}=$	1	·866	·5	0	—·5	—·866	—1	—·866	—·5	0	·5	·866
$+j$	0	·5	·866	1	·866	·5	0	—·5	—·866	—1	—·866	—·5
$S_{-2}=$	1	·5	—·5	—1	—·5	·5	1	·5	—·5	—1	—·5	·5
$+j$	0	·866	·866	0	—·866	—·866	0	·866	·866	0	—·866	—·866
$S_{-3}=$	1	0	—1	0	1	0	—1	0	1	0	—1	0
$+j$	0	1	0	—1	0	1	0	—1	0	1	0	—1
$S_{-4}=$	1	—·5	—·5	1	—·5	—·5	1	—·5	—·5	1	—·5	—·5
$+j$	0	·866	—·866	0	·866	—·866	0	·866	—·866	0	·866	—·866
$S_{-5}=$	1	—·866	·5	0	—·5	·866	—1	·866	—·5	0	·5	—·866
$+j$	0	·5	—·866	1	—·866	·5	0	—·5	·866	—1	·866	—·5
$S_{-6}=$	1	—1	1	—1	1	—1	1	—1	1	—1	1	—1
$+j$	0	0	0	0	0	0	0	0	0	0	0	0

of equation (8) and our polar plot we obtain the following values for the various sequence operators. These may be tabulated in the manner shown (see Table I.).

Suppose we are given the following curve of period T (fig. 2). We divide the interval from $t=0$ to $t=\mathrm{T}$ into twelve equal parts and the ordinates $y_0, y_1, \ldots y_{11}$ are measured.

Fig. 2.



By direct measurement from the curve, and letting $c=\mathrm{T}/12$, we construct Table II.

TABLE II.

t . .	0	c	$2c$	$3c$	$4c$	$5c$	$6c$	$7c$	$8c$	$9c$	$10c$	$11c$
u . .	9·3	15·0	17·4	23·0	37·0	31·0	15·3	4·0	$-8{\cdot}0$	$-13{\cdot}2$	$-14{\cdot}2$	6·0

From these values we form the matrix $[u]$:

$$(9)\quad [u]=\begin{Vmatrix} 9{\cdot}3 \\ 15{\cdot}0 \\ 17{\cdot}4 \\ 23{\cdot}0 \\ 37{\cdot}0 \\ 31{\cdot}0 \\ 15{\cdot}3 \\ 4{\cdot}0 \\ -\ 8{\cdot}0 \\ -13{\cdot}2 \\ -14{\cdot}2 \\ -\ 6{\cdot}0 \end{Vmatrix} \frac{1}{12} = \begin{Vmatrix} {\cdot}775 \\ 1{\cdot}250 \\ 1{\cdot}450 \\ 1{\cdot}920 \\ 3{\cdot}080 \\ 2{\cdot}580 \\ 1{\cdot}270 \\ {\cdot}333 \\ -\ {\cdot}666 \\ -1{\cdot}100 \\ -1{\cdot}180 \\ -\ {\cdot}500 \end{Vmatrix}.$$

By matrix multiplication we obtain

$$\begin{aligned}
A_0 &= S_0\,[u] = 9{\cdot}22,\\
A_{-1} &= S_{-1}[u] = -3{\cdot}45 + j\,10{\cdot}54,\\
A_{-2} &= S_{-2}[u] = 1{\cdot}725 - j\,1{\cdot}40,\\
A_{-3} &= S_{-3}[u] = 1{\cdot}65 + j\,{\cdot}99,\\
A_{-4} &= S_{-4}[u] = -{\cdot}31 + j\,{\cdot}535,\\
A_{-5} &= S_{-5}[u] = {\cdot}30 - j\,{\cdot}51,\\
A_{-6} &= S_{-6}[u] = {\cdot}125.
\end{aligned}$$

Since by (12) $A_m = \bar{A}_{-m}$ we obtain the coefficients with the positive subscripts by taking the conjugates of the various coefficients with the corresponding negative subscripts.

If it is desired to express the function in the real form of the Fourier series (16), then the coefficients a_s and b_s may be computed from the A_s coefficients by (18) and (19). In this example we thus obtain

$$\begin{aligned}
a_0 &= 9{\cdot}22, & b_1 &= 21{\cdot}08,\\
a_1 &= -6{\cdot}90, & b_2 &= -2{\cdot}80,\\
a_2 &= 3{\cdot}45, & b_3 &= 1{\cdot}98,\\
a_3 &= 3{\cdot}30, & b_4 &= 1{\cdot}07,\\
a_4 &= -{\cdot}62, & b_5 &= -1{\cdot}02.\\
a_5 &= {\cdot}60, & &\\
a_6 &= {\cdot}25. & &
\end{aligned}$$

As a rule, however, the more compact complex form of the Fourier series (1) is more convenient for calculations involving integrals or derivatives of the series. This is particularly true in electrical circuit theory where the function $u(t)$ usually plays the rôle of an electromotive force.

It may be mentioned that, exactly as in the case of the theory of Symmetrical Components, the matrix multiplication of equation (10) may be simply effected graphically directly from the polar plot of the quantities a, a^2, a^3, etc.

VII. *The Case of only Odd Harmonics Present.*

Most problems in electrical circuit theory present the analysis of waves that have the property that the second-half period is merely a repetition below the axis of the first-half period. It is then convenient to choose the axis mid-way between the highest and lowest points of the wave. It is easily shown that in such cases the Fourier series contains only the odd harmonics.

Let us consider how the process which we have developed for the general case may be modified in order to deal with this special case that is of such importance in circuit theory. To be specific, let us consider the determination of the odd harmonics up to the eleventh. To do this, let the origin be chosen at the point where the wave crosses the axis, and let the portion of the wave from $t=0$ to $t=\mathrm{T}/2$ be divided into twelve equal parts and the eleven ordinates u_1, u_2, . . . u_{11} at these points of division be measured. The series is now of the form

$$u(t)=\sum_{s=-11}^{s=+11} \mathrm{A}_s e^{2\,jst/\mathrm{T}} \quad \underline{s \text{ odd}}. \quad . \quad . \quad . \quad . \quad . \quad (20)$$

An analysis similar to that of section IV. gives the following formula for the coefficients

$$\mathrm{A}_{-k}=\mathrm{P}_{-k}[u] \quad \underline{k \text{ odd}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where P_{-k} is a line matrix similar to the sequence operator S_{-k}, but defined by the following equation:

$$\left.\begin{aligned} \mathrm{P}_{-k}&=[b^k, b^{2k}, b^{3k}, \ldots b^{11k}], \\ b&=e^{j\pi/12}=\underline{\angle 15^\circ}; \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

and $[u]$ is a columnar matrix given by

$$[u]=\left\| \begin{matrix} u_1 \\ u_2 \\ \text{—} \\ u_{11} \end{matrix} \right\| \frac{1}{12}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

The sequence operators P_{-k} are easily formed by the use of a polar plot of the vectors b, b^2, b^3, etc. The odd harmonic coefficients are then obtained by (21).

In general, if it is desired to determine the nth odd harmonic, the half-period is divided into $(n+1)$ equal intervals. The odd harmonics are then given by (21), where now

$$[u]=\left\| \begin{matrix} u_1 \\ u_2 \\ \text{—} \\ u_n \end{matrix} \right\| 1/(n+1), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

and

$$\left.\begin{aligned} \mathrm{P}_{-k}&=[b^k, b^{2k}, \ldots b^{nk},] \quad \underline{k \text{ odd}}, \\ b&=e^{j\pi/(n+1)}=\underline{\angle\left(\frac{180}{n+1}\right)^\circ}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

As before, the matrix multiplication (21) may be effected graphically if a polar plot of b, b^2, etc., is constructed.

VIII. *Conclusion.*

The natural manner in which sequence operators enter into harmonic analysis of graphically known functions has been demonstrated. Since the sequence operators may be easily computed by means of a polar plot, the determination of the various harmonic coefficients reduces to only one type of operation, that of matrix multiplication. The matrix multiplication may either be performed by a mechanical calculating machine or, if approximate and quick results are desired, it may be performed graphically. It thus appears that since there is only one type of operation involved in the method discussed, it is somewhat simpler than methods involving complicated schedules.

IX. *Bibliography.*

(1) Fortescue, C. L., "Method of Symmetrical Coordinates applied to the Solution of Polyphase Networks." Proc. A.I.E.E. xxxvii. pp. 1027–1115 (1918).

(2) Whittaker, E. T., and Robinson, G., 'The Calculus of Observations. Blackie and Son, 1924.

(3) Guillemin, E. A., 'Communication Networks,' vol. i. John Wiley and Sons, 1931.

(4) Pipes, L. A., "Matrices in Engineering." 'Electrical Engineering,' vol. lvi. pp. 1177–1190 (September, 1937).

IX. *On a Hystero-Differential Equation arising in a Probability Problem.*

By Ludwik Silberstein, Ph.D.*

[Received March 10, 1939.]

The usual differential equation is a relation between the derivatives $f'(t)$, $f''(t)$, etc. of the unknown function and the function $f(t)$ itself, all taken for the same value of the independent variable t, say, at the same instant of time t. There are situations, however, which can be described only by a relation between the derivatives at the instant t and the value of the function at some other, say past instant, $f(t-\tau)$, where τ is a fixed finite interval. The relation may also contain the past values of some derivatives of the function. Such a relation may be appropriately called a *hystero-differential* equation.

I do not know whether equations of this kind have been treated before. They are, of course, more simple, but perhaps not less interesting, than Volterra's "integro-differential" equations which, besides derivatives, contain an integral of the unknown function extended over a whole interval of the independent variable.

The probability problem † which has led me to consider such an equation may be stated, in its general form, as follows.

Let ct, where c is a constant, be the probable number of haphazard occurrences of an event during any time t. Let τ be an interval of fixed length, but placed anyhow, *e. g.*, a second. What is the probability, $P(t)$, that, in a trial of total duration t, the event will have happened twice or more than twice within a second, that is to say, within any span of time contained in t and not exceeding τ?

If $t \leqslant \tau$, the answer is, simply, $1-e^{-ct}(1+ct)$, since the probability of non-occurrence is e^{-ct} and that of a single occurrence, cte^{-ct}. If, however, $t>\tau$, the interval τ can be placed in t in an infinity of ways, and the resulting probability P becomes an intricate function of t. It then no longer contains the product ct alone, but also c separated from t, namely, through $c\tau$. (It is this feature which, in the photographic problem, constitutes a "reciprocity failure," *i. e.*, the dependence of a photographic effect, at constant total "exposure" ct, on the choice of the light intensity, c.)

---

* Communicated by the Author.

† Actually arising in the quantum theory of the photographic reciprocity failure. The "event" is then the photo-efficient absorption of a light quantum by a grain of the emulsion; t is the total exposure time at a constant intensity proportional to c, and the question is: What percentage of grains will have absorbed two or more quanta within any time not exceeding τ?

Let dP be the increment of P corresponding to a protraction of the trial time by dt. To set up the equation for P, denote by $Q(t)$ the probability that the event has happened just once between the instants $t-\tau$ and t. Then, manifestly, $dP=Q.c\,dt$, or, if we put $y=ct$ and denote differentiation with respect to y by a dash,

$$P'(y)=Q(y).$$

Now, a simple reasoning will show that, by the definition of Q,

$$Q(y)=[1-P(y)]-[1-P(y-c\tau)]e^{-c\tau},$$

the meaning of the two bracketed terms being obvious and the exponential factor of the second bracket expressing the probability of non-occurrence of the event during the interval τ expiring at t.

Thus, if we write $c\tau=\eta$, the equation for the required probability becomes

$$P'(y)=1-P(y)-[1-P(y-\eta)]e^{-\eta},\ y\geqslant\eta, \quad . \quad . \quad . \quad . \qquad (1)$$

where η is a (positive) constant. This, then, is a hystero-differential equation for $P(y)$, valid for any y not less than η *.

To solve it, put

$$P(y)=1-e^{-y}f(y). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

Then $P'=e^{-y}(f-f')$, $[1-P(y-\eta)]e^{-\eta}=e^{-y}f(y-\eta)$, and equation (1) is reduced to

$$f'(y)=f(y-\eta), \quad y\geqslant\eta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

a more simple hystero-differential equation, in fact, the simplest conceivable, which will turn out, however, to offer several points of interest.

Let us disregard, for a moment, the connexion of $f(y)$ with the probability $P(y)$ through (2) and our knowledge of $P(y)$ for the interval 0 to η. The equation (3), as it stands, is equivalent in some respects to an ordinary linear differential equation of an infinitely high order. For we may write, formally, $f(y-\eta)=f(y)-\eta f'(y)+\frac{1}{2\,!}\eta^2f''(y)-\ldots$. Thus, to determine f completely, an infinity of point data is required. Equivalently, and without assuming its developability into a Taylor series, it is manifest

---

* For $0<y<\eta$ we have, manifestly,

$$P'(y)=1-P(y)-e^{-y}.$$

that the function f can be freely prescribed within a complete interval η, say the first interval, $0 \leqslant y \leqslant \eta$, provided only that $f'(\eta)=f(0)$. The prescribed branch of f need not even be continuous. The function $f(y)$ can then be written down directly, after a simple integration, for the second interval, thence for the third interval, and so on. The freedom of prescribing a whole branch of the function being equivalent to that of prescribing an infinity of point data, the equation (3) must admit an infinity of independent particular integrals. And so it does. In fact, put $f=ae^{my}$. Then the coefficient a remains arbitrary, and the equation (3) is satisfied provided that

$$me^{\eta m}=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Now, this transcendental equation has an infinity of distinct roots m_κ, viz., one real and an infinity of pairs of conjugate complex ones. There is thus an infinity of independent integrals. There are, moreover, no double roots, as is easily ascertained. Thus the complete integral of (3) will be of the form

$$f(y)=\Sigma a_\kappa e^{m_\kappa y}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where a_κ are arbitrary constants and the summation is to be extended over all roots.

We shall return to this general aspect of the equation (3) presently. In the first place, however, let us determine the function $f(y)$ under the actual conditions of the probability problem.

If $t<\tau$ or $y<\eta$, we have $P(y)=1-e^{-y}(1+y)$, and therefore, using suffixes for the branches of the function f in the first, the second, etc. intervals (0 to η, η to 2η, etc.),

$$f_1(y)=1+y. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.1)$$

The problem consists now in the simple construction of the successive branches of the function.

If y is within the second, $y-\eta$ is within the first interval. Thus, by (3),

$$f_2'(y)=f_1(y-\eta)=1+y-\eta,$$

whence $f_2(y)=y+\frac{1}{2}(y-\eta)^2+\alpha$, where α is a constant. Since the probability P is to be continuous at $y=\eta$, so also is f. Hence, $\eta+\alpha=1+\eta$, $\alpha=1$, and

$$f_2(y)=1+y+\tfrac{1}{2}(y-\eta)^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.2)$$

It now so happens that also f', and therefore, $P'=e^{-y}(f-f')$, is continuous at $y=\eta$. (But f'' is discontinuous, jumping from 0 to 1.) By (2), for a trial time $\tau<t<2\tau$,

$$P(y)=1-e^{-y}\{1+y+\tfrac{1}{2}(y-\eta)^2\}.$$

This compact probability formula may be verified by an exhaustive enumeration of the possibilities *.

The branch f_2 being thus determined, we have, for the third interval, $f_3'(y)=f_2(y-\eta)=1+y-\eta+\frac{1}{2}(y-2\eta)^2$. Integrating and claiming the continuity of f at $y=2\eta$,

$$f_3(y)=1+y+\frac{1}{2\,!}(y-\eta)^2+\frac{1}{3\,!}(y-2\eta)^3. \quad . \quad . \quad . \quad . \quad (6.3)$$

Similarly for the fourth interval, and so on. (Notice that at $y=2\eta$ not only f', but also f'' is continuous, at $y=3\eta$, f', f'', and f''' are continuous, and so on.)

Generally, in the $(n+1)$ *st* interval,

$$f_{n+1}(y)=1+y+\frac{1}{2\,!}(y-\eta)^2+\frac{1}{3\,!}(y-2\eta)^3+\ldots+\frac{1}{(n+1)\,!}(y-n\eta)^{n+1}. \quad (6)$$

The corresponding probability is $\mathrm{P}(y)=1-e^{-y}f_{n+1}(y)$. At the same time, we see that at $y=n\eta$ not only P (by construction), but also its first n derivatives are all continuous. In other words, the successive branches of the function f or P are more and more intimately joined to each other. Each term of the polynomial (6), with the exception of the first two, is smaller than the corresponding term of the exponential series, $1+y+\frac{y^2}{2\,!}+\ldots.$, so that, for any y whatever, $f(y)<e^y$, and the more so, the greater y. In fact, when y increases indefinitely, the ratio $f(y):e^y$ tends to zero, and P to unity, for any finite η. If $\eta=\infty$, the polynomial reduces, for any finite y, to $f=1+y$, and we fall back, of course, upon the usual probability formula. If, on the other hand, the restrictive interval η becomes evanescent, the polynomial, for any finite y, becomes an infinite series, $f(y)$ tends to e^y and the probability P, to zero of course. More accurately, as follows at once from (3), we then have $f(y)=e^{y(1-\eta)}$ and, therefore, $\mathrm{P}=1-e^{-\eta y}\rightarrow\eta y$.

---

* The event can happen for the first time at any instant and then either (1) happen again at least once before the expiration of τ, or (2) fail to do so and happen at least twice in the remaining time. If we now write t_1 for the total trial time, and $y_1=ct_1$, the contribution of the first alternative to the probability P is

$$\int_0^{y_1-\eta} e^{-y}(1-e^{-\eta})\,dy+\int_{y_1-\eta}^{y_1} e^{-y}\,(1-e^{-y_1+y})\,dy,$$

and that of the second alternative,

$$e^{-\eta}\int_0^{y_1-\eta}[1-e^{-y_1+y+\eta}(1+y_1-y-\eta)]e^{-y}\,dy.$$

Performing the obvious integrations, writing again y for y_1, and adding up the two items, the reader will find $\mathrm{P}(y)=1-[1+y+\frac{1}{2}(y-\eta)^2]\,e^{-y}$, as above.

Formula (6), with (2), solves the probability problem itself, giving $P(y)$ for any desired value of y. With regard to practical application (numerical computation), the fact that $f(y)$ is thus represented separately, for each interval, by a distinct polynomial, forms no inconvenience.

Theoretically, however, it has seemed interesting to inquire whether a function $f(y)$ satisfying the equation (3), and having $f_1 = 1+y$ or any other given form of $f_1(y)$ for its first branch, can be represented by a unique expression, no matter how complicated, but valid for all, say positive y.

Since (5) is the complete integral of (3), any such expression must be reducible to an infinite series of the form (5), where everything but the coefficients a_κ is completely determined. For all m_κ, as roots of (4), may be considered as known or assignable, for any fixed numerical value of η. Suppose, then, that the function $f=f_1$ is given arbitrarily for $0 \leqslant y \leqslant \eta$, with the only restriction that it is integrable * and that

$$f_1'(\eta)=f_1(0).$$

Does that enable us to determine uniquely all the coefficients a_κ of the series (5)? The answer is in the affirmative. Although the functions $e^{m_\kappa y}$ do not form an "orthogonal" system in the sense of Fourier or (say) that of spherical harmonics, the coefficients of the terms can be isolated, one by one, in a simple way.

In fact, let $\kappa \neq \lambda$. Then, since (4) has no equal roots, $m_\kappa \neq m_\lambda$, and

$$\int_0^\eta e^{(m_\kappa - m_\lambda)y}\,dy = \frac{e^{(m_\kappa - m_\lambda)\eta}-1}{m_\kappa - m_\lambda},$$

and since, by (4), the exponential is equal to m_λ/m_κ,

$$\int_0^\eta e^{(m_\kappa - m_\lambda)y}\,dy = -\frac{1}{m_\kappa}, \quad \kappa \neq \lambda,$$

independent of m_λ, while for $\kappa=\lambda$ the integral is equal to η. Thus, if we multiply (5) by $e^{-m_\lambda y}$ and integrate from 0 to η, term by term (which presupposes the convergency of the series),

$$\int_0^\eta f(y)e^{-m_\lambda y}\,dy = \eta a_\lambda - \sum_{\kappa\neq\lambda}\frac{a_\kappa}{m_\kappa} = \left(\eta + \frac{1}{m}\right)a_\lambda - \sum \frac{a_\kappa}{m_\kappa},$$

where the last sum extends over all roots. The integral on the left hand being a known number, this is a linear equation for a_λ, containing, however, the sum in which appear all the coefficients. Apparently, then, we have an infinite system of linear equations for the coefficients, which would greatly complicate the situation. But it so happens that the sum on the right hand can be at once evaluated. In fact,

---

* For then and then only can it be continued beyond the first interval (as above) and thus can be considered as a solution of (3).

$$f(\eta)=\Sigma a e^{m\eta}=\Sigma\frac{a}{m}, \text{ by (4).}$$

Hence,

$$\left(\eta+\frac{1}{m_\kappa}\right)a_\kappa=f(\eta)+\int_0^\eta f(y)e^{-m_\kappa y}\,dy, \quad \ldots \ldots \quad (7)$$

which finds all the coefficients a_κ in terms of the function f prescribed for the interval 0 to η. Thus also $f(y)=\Sigma a_\kappa e^{m_\kappa y}$ is determined for all y.

Rigorously, we are entitled only to assert that, if the given function is at all developable into such a series, these, (7), are its coefficients. It seems likely, however, that, apart from some elaborately singular functions, any function satisfying the condition $f'(\eta)=f(0)$ can be thus developed.

To illustrate the subject in general and to derive, at the same time, some interesting properties of the roots of the equation (4), it will be enough to consider in all detail a single example.

Let us take $\eta=1$, so that the equations (3) and (4) will become $f'(y)=f(y-1)$ and

$$me^m=1. \quad \ldots \ldots \ldots \quad (8)$$

Let, as in our probability problem,

$$f(y)=1+y, \text{ for } 0\leqslant y\leqslant 1. \quad \ldots \ldots \quad (9)$$

Then

$$\int_0^1 f(y)e^{-m_\kappa y}\,dy=\frac{1}{m_\kappa^2}-2, \quad f(1)=2,$$

and, by (7),

$$a_\kappa=\frac{1}{m_\kappa(1+m_\kappa)}.$$

Thus, and omitting the obvious suffixes,

$$f(y)=\Sigma\frac{e^{my}}{m(1+m)}=\Sigma\frac{e^{m(y+1)}}{1+m} \quad \ldots \ldots \quad (10)$$

The last form of the sum will be convenient for the sequel.

Before proceeding any further we may note that, since $f(0)=1$, the equation (10) claims

$$\Sigma\frac{1}{m(1+m)}=1, \quad \ldots \ldots \ldots \quad (10^\circ)$$

a remarkable property of the roots of (8), which will be verified later on. The summation, of course, is to be extended over all roots. Similarly, from $f(1)=2$ we derive

$$\Sigma\frac{1}{m^2(1+m)}=2.$$

Let us now look for the roots of the characteristic equation (8). One, and only one, is real, say $m_0=r_0$. It is positive and less than unity. All others are complex, conjugate in pairs of course, with negative and rapidly increasing real parts.

The unique real root is readily found from (8), as it stands, *i. e.*, from

$$r_0=-\log r_0.$$

First of all, a rough sketch will show that the curve $x=-\log r$ crosses the straight line $x=r$ once only, namely, between $r=0\cdot5$ and $r=0\cdot6$. Thus, starting with 0·55 (on the right hand of the equation), one finds, as a second approximation, 0·58, and so on, with oscillations of rapidly decreasing amplitude. Repeating the process some seven or eight times, one finds, to five decimals,

$$r_0=0\cdot56714,$$

which will suffice for our purposes.

Now for the complex roots. Put $m=re^{i\theta}$, with r essentially positive. Then (8) splits into the two equations

$$r=-\frac{\theta}{\sin\theta},\quad \cos\theta=-\frac{1}{r}\log r \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The case $\theta=0$ is already covered by the root r_0. All other roots correspond to $\theta\neq0$ (and, of course, also $\theta\neq2\pi n$). If a pair r, θ satisfies both equations, so does $r, -\theta$ (giving the conjugate root). Thus it is enough to consider positive θ and (since $r>0$) only in the third quadrant, *i. e.*, of the form $\theta=(1+2n)\pi+\alpha$, where $0<\alpha<\pi/2$ and $n=0, 1, 2, 3$, etc. A graph of the two equations (11), with $r\cos\theta$ and $r\sin\theta$ as coordinates, shows at once that for each value of n there is one and only one intersection point or root m (with the corresponding conjugate one). Since $\cos\theta$ is negative, the second equation shows that $\log r$ is positive or $r>1$. But, by (8),

$$m=-\log r-i\theta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Thus the real part of every complex root is *negative*, as mentioned before. The first root is of the form $\theta=\pi+\alpha$. Start with any value of α, say 20°. Substitute $\theta=\pi+20°$ in the first of (11), finding a first approximation to r, actually $r=10\cdot20$. Substitute this in the second equation, finding a second approximation to θ, and so on. The process gives a very rapid convergence, even with a start as poor as $\alpha=20°$. In fact, the successive approximations in this case are :—

$\alpha=20°$	76°·86	70°·63	70°·68
$r=10\cdot20$	4·603	4·635	4·6363.

The last approximation may suffice for our purposes. In much the same way the second root, of the form $3\pi+\alpha$, is located, and so on. It will be enough to tabulate here the roots up to the fifth complex pair, to four significant figures.

Root.	$r.$	$\pm\theta.$
m_0	0·56714	0
m_1	4·636	$\pi+70^\circ{\cdot}68$
m_2	11·04	$3\pi+77^\circ{\cdot}44$
m_3	17·75	$5\pi+80^\circ{\cdot}68$
m_4	23·64	$7\pi+82^\circ{\cdot}31$
m_5	29·73	$9\pi+83^\circ{\cdot}45$

Manifestly, α tends to a right angle, $-\sin\theta$ to unity, so that, by the first of (11), r tends to θ itself. In fine, for the roots of high order κ we have

$$r_\kappa \doteq \theta_\kappa \doteq 2\kappa\pi-\frac{\pi}{2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

with an accuracy increasing with κ.

Let us now return to the series for $f(y)$. Of the two equivalent forms given in (10) the second is more convenient for development. Separating the term which corresponds to the real root and denoting by m, m' any pair of conjugate roots, we have

$$f(y)=\frac{e^{r_0(y+1)}}{1+r_0}+\Sigma\frac{(1+m')e^{m(y+1)}+(1+m)e^{m'(y+1)}}{(1+m)(1+m')},$$

where the sum is to be extended over all pairs, each pair of course giving, jointly, a real contribution. First of all,

$$(1+m)(1+m')=1+2r\cos\theta+r^2.$$

Again, by (12), for any m, including $m_0=r_0$,

$$e^{my}=\frac{1}{r^y}(\cos\theta y-i\sin\theta y),$$

$$e^{m'y}=\frac{1}{r^y}(\cos\theta y+i\sin\theta y).$$

Thus the series becomes, after obvious reductions,

$$f(y)=\frac{1}{(1+r_0)r_0^{\,y+1}}+2\,\Sigma\frac{(1-\log r)\cos\theta(y+1)+\theta.\sin\theta(y+1)}{(1+2r\cos\theta+r^2).r^{y+1}}, \quad . \quad (14)$$

the summation to be extended over all pairs r_κ, θ_κ from $\kappa=1$ to $\kappa=\infty$, with θ positive only. This holds for the function corresponding to

$$f(y)=1+y, \text{ for } 0\leqslant y\leqslant 1.$$

Since $r_0<1$, the first term increases with y, but all other r_κ being greater than 1, and increasing with κ, all terms under Σ decrease with increasing y.

As a matter of fact, the series converges rapidly for all (positive) y. It will be seen to represent the function $f(y)$ not only in the first interval, as prescribed, but also in all other intervals, and thus the probability function as given, piecewise, by the polynomials (6).

Let us see, in the first place, whether and how rapidly the series (14) approaches unity for $y=0$, thus verifying the property (10°) derived above for the roots m.

For $y=0$, the first term alone gives the positive contribution 1·1248. All the remaining terms are negative, and comparatively small. By means of the little table I find, for the first five of these,

$$-(0\cdot08772+0\cdot01562+0\cdot00602+0\cdot00352+0\cdot0023),$$

which brings the result down to 1·0096. The value of any of the following terms is, to all purposes, given by

$$-\frac{2\theta_\kappa}{r_\kappa^3}=-\frac{2}{r_\kappa^2}=-\frac{2}{\pi^2(2\kappa-\frac{1}{2})^2}\doteq-\frac{1}{2\pi^2\kappa^2},$$

so that the sum of the series, and thus of all $1/m(1+m)$, is

$$\Sigma\frac{1}{m(1+m)}=1\cdot0096-\frac{1}{2\pi^2}\sum_6^\infty\frac{1}{\kappa^2}.$$

Now, since $\sum_1^\infty\frac{1}{\kappa^2}=\frac{\pi^2}{6}$, the second term is 0·0092, which ought to suffice as a verification of the said property of the roots, if it needed at all a verification. The series (14) converges, at any rate, rapidly enough for $y=0$. It will, no doubt, converge more rapidly for $y>0$.

For $y=1$, the first term of the series (14) is 1·984, making up the bulk of $f(1)$. The remaining terms are now positive and smaller than in the preceding case. The first of these is 0·021, the next amounts only to 0·00095, so that the accurate value $f(1)=2$ is but slightly overshot. Thus we have also an approximate verification of the property $\Sigma\frac{1}{m^2(1+m)}=2$, if it needed one.

So far the first interval, for which f has been prescribed. For larger values of y, the first term of the series (14) becomes dominant, and it is interesting to see how nearly it represents the probability function $f(y)$, which is given accurately by (6), beyond the first interval. Now, by formula (6), with $\eta=1$, we have $f(2)=3\cdot500$, $f(3)=6\cdot167$, $f(4)=10\cdot8751$, while the corresponding values of $1:(1+r_0).r_0^{y+1}$ are

3·498, 6·168, 10·875.

Thus, the first term of the series makes up, indeed, the bulk of the probability function. The remaining terms, which contain the trigonometrical functions of $\theta(y+1)$, are now positive and now negative, making up the difference.

Similar developments, corresponding to other forms of the branch $f_1(y)$ of the function, may be left to the care of the reader. The one here discussed at some length may suffice to illustrate the properties and implications of the equation (3) and the associated characteristic equation $me^m=1$. That equation has, apart from its connexion with a probability problem, seemed particularly interesting in view of the facility with which the coefficients a_κ could be isolated. Such is by no means the case for other hystero-differential equations, even of a simple structure. Thus, *e. g.*, to mention only the next simple one, the relations in the case of the equation $f'(t)=\frac{1}{2}[f(t-\tau)+f(t+\tau)]$ are by far more intricate.

Rochester, N.Y.
February 27, 1939.

X. *Effusion Phenomena in Relativistic Quantum Statistics.*

By D. V. Gogate, M.Sc., Ph.D.(London), Physics Department,
Baroda College, Baroda (India) *.

[Received October 31, 1939.]

Introduction.

In a recent paper Srivastava (1938) has derived expressions for the number and mass of particles effusing through a narrow orifice, for an ideal gas, taking account of quantum statistics. In this paper expressions for the amount of energy effusing through such an orifice for the various cases of non-degeneracy and degeneracy are derived. The results are displayed graphically at the end to bring out clearly the physical meaning of the various formulæ involved in the treatment of effusion phenomena.

1.1. Let us consider an ideal gas enclosed in a vessel having a narrow orifice such that the mean free path of the particles is sufficiently great compared with the size of the orifice. Then, following the usual procedure (Srivastava, 1938), we get the number of particles effusing per second per unit area,

$$\mathrm{N}=\frac{\pi g}{c^2h^3}\int_0^\infty \frac{(\epsilon^2+2\epsilon mc^2)\,d\epsilon}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta}\,. \qquad (1)$$

If we multiply the above expression for the number of particles effusing by the energy ϵ, the expression for the energy effusing per second becomes

$$\mathrm{E}=\frac{\pi g}{c^2h^3}(k\mathrm{T})^4\int_0^\infty \frac{\left(u^3+\dfrac{2mc^2}{k\mathrm{T}}u^2\right)du}{1/\mathrm{A}\,.\,e^{\epsilon/k\mathrm{T}}+\beta}\,, \qquad (2)$$

where $u=\frac{\epsilon}{k\mathrm{T}}$, $\beta=+1$ for Fermi-Dirac statistics and -1 for Bose-Einstein statistics.

1.2. We shall first consider the non-degenerate case ($\mathrm{A}\ll 1$). Expanding the integrand in equation (2) in an infinite series and using the relation

$$\int_0^\infty u^n\mathrm{A}^k e^{-ku}\,du=\frac{n\,!\,\mathrm{A}^k}{k^{n+1}}\,, \qquad (3)$$

---

* Communicated by Sir Owen Richardson, F.R.S.

the expression for energy effusing per second per unit area assumes the form

$$E=\frac{\pi g}{c^2h^3}(kT)^4 . 2A\left[3+2y-\beta A\left(\frac{3}{16}+\frac{y}{4}\right)+A^2\left(\frac{1}{27}+\frac{2y}{27}\right)\right], \quad . \quad (4)$$

where $$y=\frac{mc^2}{kT}.$$

This is the general expression for the energy effusing per second per unit area, taking into account the effect of relativity mechanics. Let us now consider two limiting cases of the above expression (i.) non-relativistic when $y \gg 1$ and (ii.) relativistic when $y \ll 1$. For the non-relativistic case, we know *

$$A=A_0\left[1-\frac{15}{8y}+\frac{\beta A_0}{2^{3/2}}\dots\right], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where $$A_0=\frac{nh^3}{g(2\pi mkT)^{3/2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Hence the energy effusing in this case becomes

$$E=\frac{nkT}{2}\sqrt{\frac{8kT}{m\pi}}\left[1-\frac{3}{8y}+\frac{\beta A_0}{8}(2\sqrt{2}-1)\right]. \quad . \quad . \quad . \quad (7)$$

Expression (7) reduces to the simple formula for a classical perfect gas

$$E=\frac{nkT}{2}\sqrt{\frac{8kT}{m\pi}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

when $A_0 \rightarrow 0$ and $y \rightarrow \infty$.

Expressing the energy effusing in this case in terms of the energy density $\frac{3}{2}nkT$, we get

$$\frac{\text{Rate of effusion}}{\text{Energy density}}=\frac{1}{3}\sqrt{\frac{8kT}{m\pi}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

It may also be noted that in this case

$$\frac{\text{Number of particles effusing}}{\text{Concentration}}=\frac{1}{4}\sqrt{\frac{8kT}{m\pi}}, \quad . \quad . \quad . \quad (10)$$

and the average energy per effusing particle $=2kT$. . . (11)

Similarly, we know that for the relativistic case ($y \ll 1$)

$$A=A_0^R\left[1-y+\frac{\beta A_0^R}{8}\dots\right], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where $$A_0^R=\frac{n}{8\pi g}\left(\frac{ch}{kT}\right)^3. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Hence $$E=\frac{3}{4}nkTC\left[1-\frac{y}{3}+\beta\frac{A_0^R}{16}\dots\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

---

* For a complete list of such formulæ, see Kothari and Singh (1938, pp. 145–7).

For the completely relativistic case ($y\rightarrow 0$) and for a completely non-degenerate system, this reduces to

$$E=\frac{3}{4}nkTc. \qquad \ldots\ldots\ldots \qquad (15)$$

If we express the rate of effusion of energy given by (15) in terms of the energy density $3nkT$, we get

$$\frac{\text{Rate of effusion}}{\text{Energy density}}=\frac{1}{4}c. \qquad \ldots\ldots\ldots \qquad (16)$$

In this case it is significant to note that the ratio of the number of particles effusing to the concentration n is also $1/4\,c$, while the average energy per effusing particle is $3kT$.

2.1. Let us now investigate the effusion of energy in the case of degeneracy ($A\gg 1$). In this case the integral in equation (2) can be evaluated by using the formula due to Sommerfeld (1928),

$$\int_0^\infty \frac{d\phi(u)}{1/A.e^u+1}=\phi(u_0)+\frac{\pi^2}{6}\phi^{\text{ii}}(u_0)+\frac{7\pi^4}{360}\phi^{\text{iv}}(u_0)+\ldots, \qquad . \qquad (17)$$

where

$$u_0=\log A.$$

Thus we get

$$E_{\text{deg.}}=\frac{\pi g}{c^2h^3}(kT)^4\int_0^\infty \frac{\left(u^3+2\frac{mc^2}{kT}u^2\right)du}{1/A.\,{}^{\epsilon/kT}+1}$$

$$=\frac{\pi g}{c^2h^3}(kT)^4\frac{{u_0}^4}{4}\left[1+\frac{8}{3}\frac{mc^2}{kT}\cdot\frac{1}{u_0}\right]. \qquad \ldots\ldots \qquad (18)$$

to a first approximation, using only the first term in the right-hand side of equation (17).

Hence

$$E_{\text{deg.}}=\frac{\pi g m^4c^6}{4h^3}\left[\{(1+x^2)^{1/2}-1\}^4+\frac{8}{3}\{(1+x^2)^{1/2}-1\}^3\right], \qquad . \qquad (19)$$

where

$$x=\frac{h}{mc}\left(\frac{3n}{4\pi g}\right)^{1/3} \qquad \ldots\ldots\ldots \qquad (20)$$

and

$$u_0=\log A=\frac{mc^2}{kT}[(1+x^2)^{1/2}-1]. \qquad \ldots\ldots \qquad (21)$$

Equation (19) gives the general expression for the energy effusing in the case of a completely degenerate Fermic-Dirac gas taking account of relativity mechanics. Let us again consider the two limiting cases.

For the completely non-relativistic case ($x\rightarrow 0$) equation (19) reduces to

$$E_{\text{deg.}}=\frac{\pi g m^4c^6}{12h^3}x^6=\frac{3}{128}\frac{h^3}{\pi m^2}.n^2. \qquad \ldots\ldots \qquad (22)$$

Expressed in terms of the energy density

$$\frac{2\pi g h^2}{5m}\left(\frac{3n}{4\pi g}\right)^{5/3},$$

the rate of effusion of energy in the case of non-relativistic degeneracy is given by

$$\frac{\text{Rate of effusion of energy}}{\text{Energy density}}=\frac{5}{24}\frac{h}{m}\left(\frac{3n}{4\pi g}\right)^{1/3}. \quad . \quad . \quad . \quad . \qquad (23)$$

Also

$$\frac{\text{Number of particles effusing}}{\text{Concentration}}=\frac{3}{16}\frac{h}{m}\left(\frac{3n}{4\pi g}\right)^{1/3}, \quad . \quad . \quad . \quad . \qquad (24)$$

and the average energy per effusing particle

$$=\frac{h^2}{8\pi m}\left(\frac{4\pi g}{3}\right)^{1/3}.n^{2/3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (25)$$

For the completely relativistic case ($x\rightarrow\infty$)

$$E^{R}_{\text{deg.}}=\frac{\pi g m^4 c^6}{4h^3}.x^4=\frac{3}{16}c^2 h\left(\frac{3}{4\pi g}\right)^{1/3}.n^{4/3} \quad . \quad . \quad . \quad . \qquad (26)$$

to a first approximation.

When this is expressed in terms of the energy density

$$\pi g c h\left(\frac{3n}{4\pi g}\right)^{4/3},$$

we get

$$\frac{\text{Rate of effusion of energy}}{\text{Energy density}}=\frac{1}{4}c. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (27)$$

Similarly, the ratio

$$\frac{\text{Number of particles effusing}}{\text{Concentration}}$$

comes out in this case * to be equal to $\frac{1}{4}c$, while the average energy per effusing particle is $\frac{3}{4}\left(\frac{3n}{4\pi g}\right)^{1/3}$.

2.2. It may be worth while to note here that when we consider the general expression for the rate of effusion of mass in the case of complete degeneracy, it can be easily shown † that we obtain the same expression

---

* We may note in passing that in the case of radiation emitted by a perfectly black body, the ratio of the energy emitted per second to the energy density of radiation is also $\frac{1}{4}c\left[=\frac{u}{4}c\div u\right]$, where u is the energy density.

† In his paper (1938) Srivastava has obtained this result by working out the formulæ for the non-relativistic and relativistic cases separately. He has not deduced it from the general formula as is done above.

whether we do or do not take into account the effect of relativity mechanics. The general expression for the mass effusing can be written as

$$\mathrm{M}_{\text{deg.}}=\frac{1}{4}\frac{\pi g}{c^4h^3}(k\mathrm{T})^4\left[\frac{mc^2}{k\mathrm{T}}\{(1+x^2)^{1/2}-1\}\right]^2\times\left[\frac{mc^2}{k\mathrm{T}}\{(1+x^2)^{1/2}-1\}+\frac{2mc^2}{k\mathrm{T}}\right]^2$$

$$=\frac{\pi g}{4h^3}m^4c^4[(1+x^2)-1]^2=\frac{\pi g}{4h^3}m^4c^4x^4 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

to a first approximation.

Thus the same expression (28) holds throughout the whole region from completely non-relativistic to completely relativistic.

3.1. The simplified or asymptotic formulæ contained in the equations (8), (15), (22), (26), and (28) can be best summarized graphically. In the figure three sets of curves are drawn illustrating the effusion of energy, mass, and number of particles respectively and having for their common abscissa log n, where n is the concentration, *i. e.*, the number of particles per unit volume. The first set is a plot of log E against log n and consists of three different straight lines ED, DR, and RE′. The line ED is a graphical representation of the non-relativistic non-degenerate equation (8), while the lines DR and RE′ represent the non-relativistic degenerate (22) and relativistic degenerate (26) formulæ respectively. The lines ED and DR intersect at D, for which point A_0 is given by

$$\frac{nk\mathrm{T}}{2}\sqrt{\frac{8k\mathrm{T}}{m\pi}}=\frac{3}{128}\frac{h^3}{\pi m^2}\,.\,n^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

or

$$\mathrm{A}_0=\frac{32}{3\pi}\sim 3{\cdot}4$$

and

$$\log n=\log\left\{\frac{g\,.\,3{\cdot}4(2\pi mk\mathrm{T})^{3/2}}{h^3}\right\}=26{\cdot}7,$$

where T is assumed to be (10⁷)° K.

It will be at once noticed that for points to the left of the line log $n=26{\cdot}7$, $\mathrm{A}_0<3{\cdot}4$, and for points to its right $\mathrm{A}_0>3{\cdot}4$, and the line log $n=26{\cdot}7$ therefore separates the non-degenerate and degenerate regions.

Similarly, the lines DR and RE′ intersect at R, for which point x is given by

$$\frac{3}{16}hc^2\left(\frac{3}{4\pi g}\right)^{1/3}n^{4/3}\doteq\frac{3}{128}\frac{h^3}{\pi m^2}\,.\,n^2, \quad . \quad . \quad . \quad . \quad . \quad (30)$$

or $x=1{\cdot}7$ and log $n=30{\cdot}5$.

For points to the left of the line log $n=30{\cdot}5$, $x<1{\cdot}7$, and for points to its right $x>1{\cdot}7$, and therefore the line log $n=30{\cdot}5$ separates the non-relativistic and relativistic regions. These "critical values" of A_0 and x for the different cases are tabulated as shown :—

TABLE I.

Serial no.	Effusion of	Set of curves.	A_0.	x.
1	Energy.	ED, DR, RE′	3·4	1·7
2	Mass.	MD′, D′R, RE′	2·6	—
3	Number.	ND″, D″R″, R″N′	2·6	1·3

The second set of curves is plotted with log n as abscissa and log $M.c^2$ as the ordinate where M is the mass effusing per second per unit area. This consists of the three straight lines MD′, D′R, and RM′ respectively. The points D′ and R for this set of curves can be determined by the method explained above.

It may be noted, however, that the line RM′ is a continuation of the line D′R and coincides with the line RE′ of the first set (the points M′ and E′ overlapping on each other). That this should be so is obvious, because (*a*) the formulæ for the mass effusing in the non-relativistic degenerate and relativistic degenerate cases respectively are identical with each other [see equation (28) above], and (*b*) in plotting the graphs of this set we have taken Mc^2 instead of M as the unit of mass effusing.

The third set of curves is a plot of log N against log n, where N is the number of particles effusing per second per unit area. This set again consists of three different straight lines ND″, D″R″, and R″N′ respectively, the points D″ and R″ being determined as before. The formulæ for the number of particles effusing in the three different cases (viz., non-relativistic non-degenerate, non-relativisitc degenerate, and relativistic degenerate) which are represented by the three straight lines ND″, D″R″, and R″N′ respectively have already been derived by Srivastava (1938).

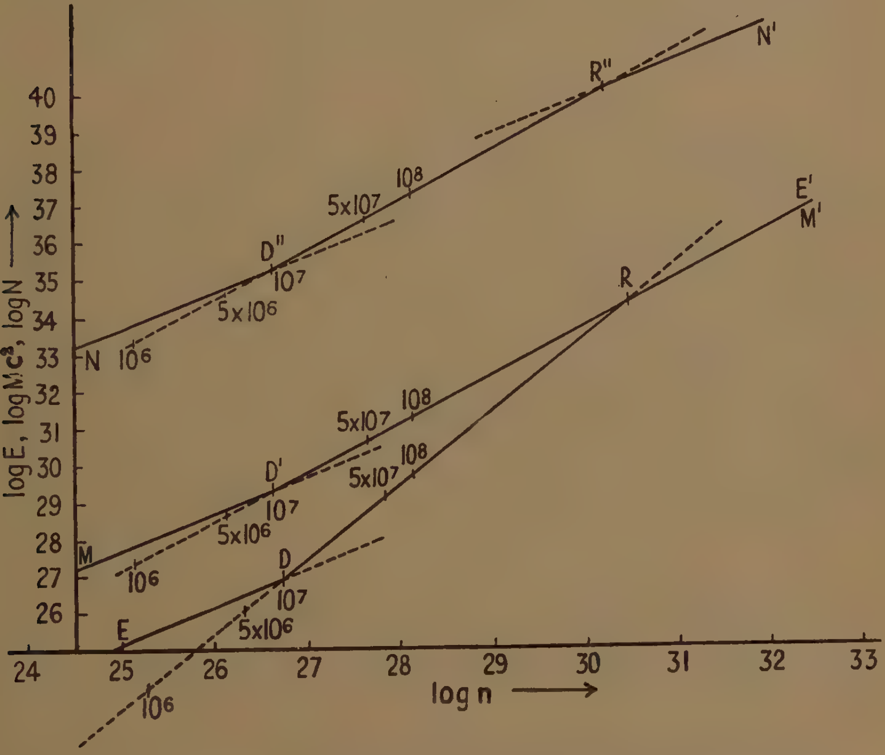
3.2. All the curves given in the figure correspond to the temperature $(10^7)°$ K. But similar curves for any other temperature can be easily drawn by constructing a ladder of points on the curves of the figure. For instance, the point D on the curve DR corresponds to $A_0=3{\cdot}4$ and

$$\log n=\log\left\{\frac{g.3{\cdot}4(2\pi mk)^{3/2}}{h^3}\right\}+\frac{3}{2}\log T,$$

where $T=(10^7)°$ K.

If we substitute different values of temperature for T in the above equation, viz., $T=10^6, 5\times10^6, 10^7, 5\times10^7$, etc., we shall get different values of log n, viz., log n=25·2, 26·2, 27·7, 28·2, etc. If now straight lines are drawn through the points given by the above values of log n and perpendicular to the abscissa, they will intersect the curve DR or its prolongation in different points corresponding to the temperatures marked against them, and the straight lines drawn through these points parallel to ED will represent the non-relativistic non-degenerate curves for those temperatures, the degenerate curves, of course, remaining unchanged.

In the figure every set of curves is provided with one ladder of such points each labelled with the corresponding value of the temperature.



Effusion of energy E, mass M ($\times c^2$) and number N plotted against concentration n (logarithms in each case).

From an inspection of the figure it is at once seen that :—

(i.) In the case of a degenerate gas, the effusion (of number, mass, or energy as the case may be) is greater than in the case of a non-degenerate (classical) gas at the same temperature.

(ii.) The effusion (of number, mass, or energy as the case may be) of a degenerate gas is very large compared with what would be computed for a " classical gas " at the same temperature and concentration.

(iii.) The effusion of a " classical gas " is very large compared with what would be computed for a degenerate gas at the same temperature and concentration.

Finally, it may be noted that all the curves in the non-relativistic non-degenerate region run parallel to one another.

In conclusion, the author wishes to express his sincere thanks to Prof. Sir O. W. Richardson, F.R.S., for his kind interest in this work, and to Dr. D. S. Kothari for a discussion of the problem and hospitality at the Delhi University.

SUMMARY.

This paper deals with the phenomenon of the effusion of energy of an ideal gas through a narrow orifice taking account of relativistic quantum statistics. In section (1) non-degenerate matter is considered and the amount of energy effusing per second per unit area is calculated. In section (2) the effusion of energy is calculated for a completely degenerate gas, degenerate in the sense of Fermi-Dirac statistics. It is also shown that, in the case of complete degeneracy, the same expression holds for the rate of effusion of mass throughout the whole region from completely non-relativistic to completely relativistic. The results are displayed graphically in section (3) to being out clearly the physical significance of the different formulæ for the various cases of non-degeneracy and degeneracy.

REFERENCES.

Srivastava, B. N. Proc. Roy. Soc. A, clxvii. p. 516 (1938).
Kothari, D. S., and Singh, B. N., *Z. Astrophys.* xv. p. 145 (1938).
Sommerfeld, A., *Z. Phys.* xlvii. p. 1 (1928).

XI. *On some Problems in the Diffraction of Heat.*

By Dr. Arnold N. Lowan †.

[Received April 30, 1939.]

Abstract.

The object of this paper is to determine the temperature distribution in a semi-infinite solid whose boundary plane is impervious to heat except for an aperture of given shape, when the solid is subjected to an influx of heat of given rate. Rigorous solutions are obtained for the following cases :—(A) The aperture is an infinite slit between two parallel lines. (B) The aperture is a circle of radius R. (C) The aperture is of finite extent and has any arbitrary shape. (D) The aperture consists of the half plane $x=0$, $y>0$. The method consists in first obtaining the Laplace Transforms of the temperature function and subsequently performing the inversion of the Laplace Transforms by means of standard methods in Operational Calculus. The above problem is the thermal equivalent of the problem of diffraction of electro-magnetic waves. For further elucidation of the mathematical technique, the reader is referred to the following papers by the author : "On the Cooling of a Radioactive Sphere," 'Physical Review,' November 1933, and the papers on "Heat Conduction and Wave Motion" by the author, published since 1933 in the 'Journal of the American Mathematical Society,' 'Bulletin of the American Mathematical Society,' and the 'Philosophical Magazine.'

The boundary plane $x=0$ of a semi-infinite solid $x>0$, is impervious to heat, except for an aperture of given shape. If the semi-infinite solid is subjected to an influx of heat at the rate Q(P, t) per unit area (where P is a point in the aperture), what is the distribution of temperatures at time t if at time $t=0$ the semi-infinite solid is at 0° C. ? The above problem is the thermal equivalent of the problem of the diffraction of light.

The object of this paper is to derive rigorous solutions for the following cases : (A) The aperture is an infinite slit between two parallel lines at distance 2 a. (B) The aperture is a circle of radius R. (C) The aperture is of finite extent and has any arbitrary shape. (D) The aperture consists of the half plane $x=0$, $y>0$.

---

† Communicated by the Author.

Case A.

The problem is evidently two-dimensional. Furthermore, if the aperture extends from $y=-a$ to $y=a$, the solution must be symmetrical with respect to the OX-axis, and therefore the temperature gradient must vanish for $y=0$. Accordingly, it suffices to define the temperature function for the domain $x>0$, $y>0$.

The mathematical formulation of the problem is as follows :

$$\left[\frac{\partial}{\partial t}-k\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)\right]\mathrm{T}(x, y\,;\,t)=0. \quad x>0,\ y>0,\ t>0. \quad . \quad . \quad (1)$$

$$\lim_{t\to 0}\mathrm{T}(x, y\,;\,t)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\frac{\partial}{\partial y}\mathrm{T}(x, y\,;\,t)=0, \quad \text{for } y=0. \quad . \quad . \quad . \quad . \quad (3)$$

$$\frac{\partial}{\partial x}\mathrm{T}(x, y\,;\,t)=\phi(y, t)=\left\{\begin{matrix}-\frac{1}{\mathrm{K}}\mathrm{Q}(y, t) \text{ inside the slit}\\ 0 \qquad\qquad \text{outside the slit}\end{matrix}\right\} \text{for } x=0, \quad (4)$$

where K is the thermal conductivity of the solid.

The Laplace Transform

$$\mathrm{T}^*(x, y\,;\,p)=\int_0^\infty e^{-pt}\mathrm{T}(x, y\,;\,t)\,dt \quad . \quad . \quad . \quad . \quad . \quad (5)$$

(where p is a complex variable whose real part is positive), of the temperature function, must satisfy the differential equation

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}-\frac{p}{k}\right)\mathrm{T}^*(x, y\,;\,p)=0, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

and the boundary conditions

$$\frac{\partial}{\partial y}\mathrm{T}^*(x, y\,;\,p)=0, \quad \text{for } y=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

$$\frac{\partial}{\partial x}\mathrm{T}^*(x, y\,;\,p)=\phi^*(y, p), \quad \text{for } x=0, \quad . \quad . \quad . \quad . \quad (8)$$

where $\phi^*(y, p)$ is the Laplace Transform of $\phi(y, t)$.

In view of the integral representation

$$\mathrm{F}(y)=\frac{2}{\pi}\int_0^\infty d\xi\int_0^\infty \mathrm{F}(\eta)\cos\xi y\cos\xi\eta\,d\eta \quad . \quad . \quad . \quad . \quad . \quad (9)$$

for a function satisfying the condition $\mathrm{F}(-y)=\mathrm{F}(y)$ it is easily seen that the expression

$$\begin{aligned}\mathrm{T}^*(x, y : p)&=\frac{2}{\pi}\int_0^\infty d\xi\int_0^\infty -\frac{e^{-\lambda x}}{\lambda}\phi^*(\eta, p)\cos\xi y\cos\xi\eta\,d\eta\\ &=\frac{2}{\pi\mathrm{K}}\int_0^\infty d\xi\int_0^a \frac{e^{-\lambda x}}{\lambda}\mathrm{Q}^*(\eta, p)\cos\xi y\cos\xi\eta\,d\eta \quad . \quad . \quad (10)\end{aligned}$$

is a solution of the differential equation (6) satisfying the boundary conditions (7) and (8) provided

$$\lambda^2-\xi^2=\frac{p}{k}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

If $u(x, \tau ; t)$ is a solution of the integral equation

$$\int_0^\infty e^{-pt}u(x, \xi ;\ t)\, dt=\frac{e^{-\lambda x}}{\lambda}=\sqrt{k}\,.\frac{e^{-\frac{x}{\sqrt{k}}\cdot\sqrt{p+k\xi^2}}}{\sqrt{p+k\xi^2}}, \quad . \quad . \quad . \quad (12)$$

then with the aid of "Borel's Theorem" we obtain the inverse Laplace Transform of (10) in the form

$$\mathrm{T}(x, y ;\ t)=\frac{2}{\pi\mathrm{K}}\int_0^t d\tau\int_0^\infty d\xi\int_0^a u(x, \xi ;\ t-\tau)\mathrm{Q}(\eta, \tau) \cos \xi y \cos \xi\eta\, d\eta. \quad (13)$$

The explicit expression for $u(x, \xi ; t)$ may be obtained as follows: In the well-known identity †

$$\int_0^\infty e^{-pt}\frac{e^{-\frac{\lambda}{t}}}{\sqrt{\pi t}}dt=\frac{e^{-2\sqrt{\lambda}\cdot\sqrt{p}}}{\sqrt{p}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

put $2\sqrt{\lambda}=\dfrac{x}{\sqrt{k}}$ and replace p by $p+k\xi^2$. This yields the identity

$$\int_0^\infty e^{-pt}.e^{-k\xi^2t}.\frac{e^{-\frac{x^2}{4kt}}}{\sqrt{\pi t}}dt=\frac{e^{-\frac{x}{\sqrt{k}}\cdot\sqrt{p+k\xi^2}}}{\sqrt{p+k\xi^2}}. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Comparison between (15) and (12) yields

$$u(x, \xi ;\ t)=\frac{\sqrt{k}}{\sqrt{\pi t}}\cdot e^{-k\xi^2t-\frac{x^2}{4kt}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

In view of (16), the solution (13) becomes

$$\mathrm{T}(x, y ;\ t)=\frac{2\sqrt{k}}{\pi^{3/2}\mathrm{K}}\int_0^t\frac{e^{-\frac{x^2}{4k(t-\tau)}}}{\sqrt{t-\tau}}\,d\tau\int_0^a \mathrm{Q}(\eta, \tau)\,d\eta$$

$$.\int_0^\infty e^{-k^2\xi(t-\tau)} \cos \xi y \cos \xi\eta\, d\xi. \quad (17)$$

The identity ‡

$$\int_0^\infty e^{-a^2\xi^2} \cos \alpha\xi\, d\xi=\frac{\sqrt{\pi}}{2a}e^{-\frac{\alpha^2}{4a^2}} \quad . \quad . \quad . \quad . \quad . \quad (18)$$

---

† J. B. Carson, 'Electric Circuit Theory and Operational Calculus,' p. 39 (1926).

‡ H. S. Carslaw, 'The Conduction of Heat,' Art. 82 (1921).

yields

$$\int_0^\infty e^{-k(t-\tau)\xi^2}\cos\alpha\xi\,d\xi=\frac{\sqrt{\pi}}{2\sqrt{k(t-\tau)}}\cdot e^{-\frac{\alpha^2}{4k(t-\tau)}}. \quad . \quad . \quad . \qquad (18')$$

If the last integral of the second member of (17) is resolved into two integrals by means of the relation

$$2\cos\xi y\cos\xi\eta=\cos\xi(y+\eta)+\cos\xi(y-\eta)$$

and we make use of (18′), (17) becomes

$$\mathrm{T}(x,\,y\,;\;t)=\frac{1}{2\pi\mathrm{K}}\int_0\frac{e^{-\frac{x^2}{4k(t-\tau)}}}{t-\tau}\left[e^{-\frac{(y+a)^2}{4k(t-\tau)}}+e^{-\frac{(y-a)^2}{4k(t-\tau)}}\right]d\tau\int_0^a\mathrm{Q}(\eta,\,\tau)\,d\eta. \qquad (19)$$

This is the final solution of our problem.

If in (19) we let a approach 0, and put the condition

$$\operatorname*{Lim}_{a\to 0}\int^a\mathrm{Q}(\eta,\,t)\,d\eta=\frac{1}{2}\mathrm{Q}_0(t),$$

we obtain the solution

$$\mathrm{T}_0(x,\,y\,;\;t)=\frac{1}{2\pi\mathrm{K}}\int_0^t\frac{e^{-\frac{x^2+y^2}{4k(t-\tau)}}}{t-\tau}\mathrm{Q}_0(\tau)\,d\tau. \quad . \quad . \quad . \quad . \qquad (20)$$

In (20) we recognize the known solution of the problem in heat conduction in a semi-infinite solid $x>0$ due to a continuous line-source along the OZ-axis when the plane $x=0$ is impervious to heat †.

Case B.

In this case, assuming that the heat influx is a function of r and t, it is obvious that the problem has cylindrical symmetry and therefore the temperature function must satisfy the system

$$\left[\frac{\partial}{\partial t}-k\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial x^2}\right)\right]\mathrm{T}(r,\,x\,;\;t)=0. \quad . \quad . \quad . \quad . \qquad (1)$$

$$\operatorname*{Lim}_{t\to 0}\mathrm{T}(r,\,x\,;\;t)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

$$\frac{\partial}{\partial x}\mathrm{T}(r,\,x\,;\;t)=\phi(r,\,t)=\left\{\begin{array}{ll}-\frac{1}{\mathrm{K}}\mathrm{Q}(y,\,t) & \text{inside the slit}\\ 0 & \text{outside the slit}\end{array}\right\}\text{for } x=0. \qquad (3)$$

The Laplace Transform $\mathrm{T}^*(r,\,x\,;\,p)$ of $\mathrm{T}(r,\,x\,;\,t)$ must satisfy the system

$$\left[\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{\partial^2}{\partial x^2}-\frac{p}{k}\right]\mathrm{T}^*(r,\,x\,;\;p)=0. \quad . \quad . \quad . \quad . \qquad (4)$$

$$\frac{\partial}{\partial x}\mathrm{T}^*(r,\,x\,;\;p)=\phi^*(r,\,p)\quad\text{for } x=0. \quad . \quad . \quad . \quad . \quad . \qquad (5)$$

---

† See H. S. Carslaw, Articles 86 and 72.

The expression

$$\int_0^\infty \xi\, d\xi \int_0^R -\frac{e^{-\lambda x}}{\lambda}\phi^*(\rho, p).\mathrm{J}_0(\xi\rho).\mathrm{J}_0(\xi r).\rho\, d\rho \quad . \quad . \quad . \quad (6)$$

is a solution of (4) satisfying the boundary condition (6) provided $\lambda^2-\xi^2=p/k$.

Introducing the function $u(x, \xi\,; t)$ defined in Case A, the inversion of (6) yields

$$\mathrm{T}(r, x\,;\, t)=\frac{1}{\mathrm{K}}\int_0^t d\tau \int_0^\infty \xi\, d\xi \int_0^R u(x, \xi\,;\, t-\tau)\,\mathrm{Q}(\rho, \tau)\,\mathrm{J}_0(\xi\rho)\ \mathrm{J}_0(\xi r).\rho\, d\rho$$

$$=\frac{1}{\mathrm{K}}\sqrt{\frac{k}{\pi}}\int_0^t \frac{e^{-\frac{x^2}{4k(t-\tau)}}}{\sqrt{t-\tau}}d\tau \int_0^R \rho\,\mathrm{Q}(\rho, \tau)\,d\rho \int_0^\infty \xi e^{-k\xi^2(t-\tau)}\mathrm{J}_0(\xi\rho).\mathrm{J}_0(\xi r)\,d\xi. \quad (7)$$

The last integral may be written in a more convenient form.

With the aid of the identity (H. S. Carslaw, Article 90)

$$\int_0^\infty e^{-p^2\xi^2}\mathrm{J}_0(a\xi)\,\mathrm{J}_0(b\xi)\,\xi\, d\xi=\frac{1}{2p^2}e^{-\frac{a^2+p^2}{4p^2}}\,\mathrm{I}_0\left(\frac{ab}{2p^2}\right). \quad . \quad . \quad . \quad (8)$$

(7) becomes

$$\mathrm{T}(r, x\,;\, t)=\frac{1}{2\mathrm{K}}\sqrt{\frac{1}{\pi k}}\int^t \frac{e^{-\frac{x^2+r^2}{4k(t-\tau)}}}{(t-\tau)^{3/2}}d\tau \int_0^R \rho e^{-\frac{\rho^2}{4k(t-\tau)}}\,\mathrm{Q}(\rho, \tau)\mathrm{I}_0\left(\frac{r\rho}{2k(t-\tau)}\right)d\rho. \quad . \quad . \quad . \quad (9)$$

If in (7) we let R approach 0 and put the condition

$$\lim_{R\to 0} 2\pi\int_0^R \rho\,\mathrm{Q}(\rho, t)\,d\rho=\mathrm{Q}_0(t), \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

we get

$$\mathrm{T}_0(r, x\,;\, t)=\frac{1}{4\pi^{3/2}\mathrm{K}\sqrt{k}}\int_0^t \frac{e^{-\frac{x^2+r^2}{4k(t-\tau)}}}{(t-\tau)^{3/2}}d\tau. \quad . \quad . \quad . \quad . \quad (11)$$

In (11) we recognize the known solution of the problem in heat conduction in a semi-infinite solid $x>0$ due to a continuous point-source at the origin when the boundary plane $x=0$ is impervious to heat.

Case C.

In this case the temperature function must satisfy the system

$$\left[\frac{\partial}{\partial t}-k\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}\right)\right]\mathrm{T}^*(x, y, z\,;\, t)=0, \quad . \quad . \quad . \quad (1)$$

$$\lim_{t\to 0}\mathrm{T}(x, y, z\,;\, t)=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\frac{\partial}{\partial z}\mathrm{T}(x, y, z\,;\, t)=\phi(x, y\,;\, t)=\left\{\begin{array}{ll}-\frac{1}{\mathrm{K}}\mathrm{Q}(x, y\,;\, t) & \text{inside the slit} \\ 0 & \text{outside the slit}\end{array}\right\}\text{ for } z=0. \quad . \quad . \quad . \quad (3)$$

The Laplace Transform of the temperature function must then satisfy the equations

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{p}{k}\right]T^*(x, y, z;\ t)=0. \quad \ldots \quad (4)$$

$$\frac{\partial}{\partial z}T^*(x, y, z;\ t)=\phi^*(x, y;\ t) \quad \text{for } z=0, \quad \ldots \quad (5)$$

where $\phi^*(x, y;\ p)$ is the Laplace Transform of $\phi(x, y;\ t)$

In view of the identity

$$F(x, y)=\frac{1}{4\pi^2}\iiiint\limits_{-\infty}^{\infty} F(\alpha, \beta)\cos\left[\xi(x-\alpha)+\eta(y-\beta)\right]d\xi\, d\eta\, d\alpha\, d\beta, \quad . \quad (6)$$

the expression

$$T^*(x, y, z;\ p)=\frac{1}{4\pi^2 K}\iint\limits_{-\infty}^{\infty} d\alpha\, d\beta \iint\limits_{S}\frac{e^{-\lambda z}}{\lambda}Q^*(\alpha, \beta)\cos\left[\xi(x-\alpha)+\eta(y-\beta)\right]d\xi\, d\eta \quad \ldots \quad (7)$$

is a solution of (4) satisfying the boundary condition (5) provided $\lambda-\xi^2-\eta^2=p/k$, where S is the region inside the aperture.

The inversion of (7) leads to the solution

$$T(x, y, z;\ t)=\frac{\sqrt{k}}{4\pi^2 K\sqrt{\pi}}\int_0^t \frac{e^{-\frac{z^2}{4k(t-\tau)}}}{\sqrt{t-\tau}}d\tau \iint\limits_{S} Q(\alpha, \beta, \tau)\, d\alpha\, d\beta$$

$$\times \iint\limits_{-\infty}^{\infty} e^{-k(\xi^2+\eta^2)(t-\tau)}\cos[\xi(x-\alpha)+\eta(y-\beta)\, d\xi\, d\eta. \quad (8)$$

With the aid of the identity (18′—A) the last double integral in (8) may be easily seen to be equal to

$$\frac{\pi}{k(t-\tau)}e^{-\frac{(x-\alpha)^2+(y-\beta)^2}{4k(t-\tau)}}.$$

The solution (8) therefore becomes

$$T(x, y, z;\ t)=\frac{1}{4\pi^{3/2}\sqrt{k}K}\int_0^t \frac{e^{-\frac{z^2}{4k(t-\tau)}}}{(t-\tau)^{3/2}}d\tau \iint\limits_{S} Q(\alpha, \beta, \tau)\, e^{-\frac{(x-\alpha)^2+(y-\beta)^2}{4k(t-\tau)}}d\alpha\, d\beta \quad \ldots \quad (9)$$

If in (9) we let S approach 0 and we put the condition

$$\lim_{s\to 0}\iint\limits_{S} Q(\alpha, \beta;\ t)\, d\alpha\, d\beta=Q_0(t), \quad \ldots \quad (10)$$

we obtain once more the known solution of the problem of heat conduction in the semi-infinite solid $z>0$ due to a continuous point-source at the origin when the plane $z=0$ is impervious to heat.

Case D.

In this case the domain of definition of the temperature function is $x>0$, $-\infty<y<\infty$. The differential equation is the same as in Case A and the boundary condition is

$$\frac{\partial}{\partial x}\mathrm{T}(x, y;\ t)=\phi(y, t)=\left\{\begin{array}{ll}-\frac{1}{\mathrm{K}}\mathrm{Q}(y, t) & \text{for } y>0\\ 0 & \text{for } y<0\end{array}\right\} \text{for } x=0. \quad . \quad (1)$$

In view of the identity

$$\mathrm{F}(y)=\frac{1}{\pi}\int_0^\infty d\xi\int_{-\infty}^\infty \mathrm{F}(\eta)\cos\xi(y-\eta)\,d\eta, \quad . \quad . \quad . \quad . \quad (2)$$

the expression

$$\mathrm{T}^*(x, y;\ p)=\frac{1}{\pi}\int_0^\infty d\xi\int_{-\infty}^\infty -\frac{e^{-\lambda x}}{\lambda}\phi^*(\eta, p)\cos\xi\{y-\eta)\,d\eta \quad . \quad . \quad (3)$$

is a solution of the differential equation (1—A), satisfying the boundary condition obtained from (1) by replacing $\mathrm{T}(x, y;\ t)$ and $\phi(y, t)$ by $\mathrm{T}^*(x, y;\ p)$ and $\phi^*(y, p)$ respectively.

Proceeding in exactly the same manner as in Case A, the inversion of (3) yields

$$\mathrm{T}(x, y;\ t)=\frac{\sqrt{k}}{\mathrm{K}\pi^{3/2}}\int_0^t\frac{e^{-\frac{x^2}{4k(t-\tau)}}}{\sqrt{t-\tau}}d\tau\int_0^\infty \mathrm{Q}(\eta, \tau)\,d\eta\int_0^\infty e^{-k\xi^2(t-\tau)}\cos\xi(y-\eta)\,d\xi. \quad . \quad . \quad . \quad (4)$$

Making once more use of the identity (18′—A) the last equation becomes

$$\mathrm{T}(x, y;\ t)=\frac{1}{2\mathrm{K}\sqrt{\pi}}\int_0^t\frac{e^{-\frac{x^2}{4k(t-\tau)}}}{(t-\tau)}d\tau\int_0^\infty e^{-\frac{(y-\eta)^2}{4k(t-\tau)}}\mathrm{Q}(\eta.\ \tau)\,d\eta. \quad . \quad (5)$$

From (5) it is apparent that $\mathrm{T}\to0$ as $y\to-\infty$ and that, furthermore,

$$\frac{\partial\mathrm{T}}{\partial y}\to0 \text{ as } y\to\infty.$$

These results may, of course, be anticipated from physical considerations.

XII. *Error in Paper on "Certain Vibration Problems solved by means of an Analogous Problem in Heat Conduction,"* Phil. Mag. xxii. p. 1079.

To the Editors of the Philosophical Magazine.

Gentlemen,—

In a short paper entitled "Note on a Problem in Heat Conduction," Phil. Mag., June 1939, pp. 706–8, Mr. Harold W. Woolley has pointed out an error in the solution given of the heat problem proposed in the above paper—the error being the omission of a term in the treatment of the problem by the wave-train method. I am very glad indeed that the error has been detected, and that the discrepancy between the solutions obtained by the two different methods used in the paper has thereby been removed.

The error made does not affect the mathematical analogies existing between the problem in heat conduction and certain vibration problems which it was the purpose of the paper to demonstrate. In the cases considered, instantaneous initial heat distribution (mass × specific heat × temperature) in the heat problem, corresponds with instantaneous initial (mass × displacement) at any specified point in the elastic problem, and the wave-train solutions representing similar initial conditions in the two systems must therefore also closely correspond.

Equation (27) in the paper is incorrect, as Mr. Woolley has shown, as the complete solution of the heat conduction problem proposed. It leads also to an incorrect solution of the corresponding elastic problem in the equivalent equation (53). Expression (33) in the paper gives the correct solution of the heat problem, and it provides also the correct solutions of the analogous elastic problems referred to in the paper. The wave-system developed in (27) satisfies initial conditions referring to the rod alone in the heat problem; and the paper demonstrates that an identical system of wave-trains represents the corresponding part of the solution of the analogous elastic problems. The error contained in the original paper is the omission of the effect of the mass M as an initial instantaneous heat source, and the omission of the corresponding term in the analogous elastic problems. The following investigation shows how the effect of the mass M as an initial instantaneous heat source can be determined by the method of wave-trains. The notation is that used in the original paper.

Let the mass M, at $x=a$, be regarded as a periodic heat source of strength qe^{ikt}, and let

$$\theta=\rho_1 e^{ikt-i\lambda(a-x)}, \quad \sqrt{\frac{ik}{\kappa}}=i\lambda, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

represent a negative wave-train issuing from the mass M as source. This train with its continuations forms the wave-system—

$$\theta=\frac{\rho_1 e^{ikt-i\lambda a}\,.\,2i\sin\lambda x}{1+\mathrm{B}e^{-i\lambda.2a}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

as explained on pp. 1081–2 of the original paper. From equation (22) of the paper we also have

$$1+\mathrm{B}e^{-i\lambda.2a}=\frac{2ie^{-i\lambda a}[\mathrm{KA}\cos\lambda a-\mathrm{M}s\kappa\lambda\sin\lambda a]}{i\mathrm{KA}-\mathrm{M}s\kappa\lambda},$$

so that (2) above becomes

$$\theta=\frac{\rho_1(i\mathrm{KA}-\mathrm{M}s\kappa\lambda)e^{ikt}\sin\lambda x}{\mathrm{KA}\cos\lambda a-\mathrm{M}s\kappa\lambda\sin\lambda a}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This solution represents a wave-system issuing from the mass M as source; and the condition that the mass M is a source of strength qe^{ikt} is that, at $x=a$,

$$\mathrm{M}s\frac{d\theta}{dt}+\mathrm{KA}\frac{d\theta}{dx}=qe^{ikt}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The wave-system (3) satisfies this condition provided that we choose q/λ as the value of $\rho_1(i\mathrm{KA}-\mathrm{M}s\kappa\lambda)$. Hence the periodic heat source solution representing the effect of a periodic source of strength qe^{ikt} at the mass M is

$$\theta=\frac{qe^{ikt}\sin\lambda x}{\lambda[\mathrm{KA}\cos\lambda a-\mathrm{M}s\kappa\lambda\sin\lambda a]}. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The corresponding instantaneous source solution representing a source of strength q is given by

$$\theta_1=\frac{1}{2\pi}\int_{-\infty}^{+\infty}dk\theta, \text{ where } \theta \text{ is given by (5).}$$

Thus
$$\theta_1=\frac{iq\kappa}{\pi}\int_{-\infty}^{+\infty}\frac{d\lambda\, e^{-\kappa\lambda^2t}\sin\lambda x}{(\mathrm{KA}\cos\lambda a-\mathrm{M}s\kappa\lambda\sin\lambda a)}$$

$$=2q\kappa\sum_1^{\infty}e^{-\kappa\lambda^2t}\sin\lambda x\sin\lambda a\,.\,\frac{\mathrm{A}_n}{\mathrm{KA}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where A_n is the coefficient defined by (26) of the original paper, and the summation refers to all positive roots of (24).

To connect this solution with the problem treated in Phil. Mag. xxii. we identify q with $\mathrm{M}sf(a)$—the initial heat content of the mass M as required by equation (6) of the paper. The expression (6) obtained above then falls to be added to the right-hand side of (27) of the paper, and (27) then comes into complete agreement with (33).

The term corresponding to (6) above in the analogous elastic problems can be found by an exactly similar investigation, the mass M being in this case an instantaneous source of (mass-displacement.) The analogy between the heat problem and the elastic problems is, in fact, complete both with respect to the rod, as shown in the original paper, and with respect to the ball and the vibrating mass.

Starting with wave-train (1), in which $\lambda = k/\mathrm{V}$, we find that u_1, the instantaneous integral required, is obtainable from (6) above by inserting corresponding terms; e. g., e^{ikt} for $e^{-\kappa\lambda^2 t}$, EA for KA, $\mathrm{MV}^2\lambda$ for $\mathrm{M}s\kappa\lambda$. It is worth noting that the ratio $(\mathrm{M}s\kappa/\mathrm{KA})$, is equal to the ratio (Heat capacity of M/Heat capacity of unit length of the rod). The corresponding ratio $(\mathrm{MV}^2/\mathrm{EA})$ is equal to the ratio (Mass M/Mass per unit length of the rod).

Yours faithfully,
G. Green.

XIII. *Notices respecting New Books.*

Rutherford: being the Life and Letters of the Rt. Hon. Lord Rutherford, O.M. By A. S. Eve. [Pp. xvi+451.] (Cambridge University Press, 1939. Price 21*s.* net.)

Rutherford's reputation as a great experimental genius and leader of science is so firmly established and so widely famed that any biography of him is bound to be of value and interest to physicists, and to receive a very general welcome. This would be true even of a biography which gave no more than a moderately competent account of his scientific work, and we need hardly say here that Professor Eve's treatment of his subject is far more than merely competent.

There was, moreover, in Rutherford, quite apart from his outstanding contributions to science, much that is well worth recording—for example, his genius for happy friendship, his immense vitality, his fresh and almost boyish enthusiasms. He had none of the tiresome eccentricities which sometimes accompany genius, and which more frequently are cultivated by those who merely aspire to a reputation for genius. In brief, Rutherford was in very many ways an extraordinarily fine and vigorous specimen of what we should like to regard as the ordinary man; one with hosts of friends to whom he delighted to talk, and to whom he perhaps just a little less eagerly listened; one with no real enemies, but perhaps a sufficiency of antipathies to give spice to his gossip; one who could at times irritate his friends, but never lose their affection. Here again, as a personal friend of very long standing, as a former colleague at McGill, and as a successor in the Macdonald Chair of Physics, Professor Eve was the obvious choice as Rutherford's biographer.

He has done his work very well indeed. In his Preface he claims modestly only to have attempted to allow Rutherford to reveal himself, in selections from his papers, books, letters, material preserved by Lady Rutherford, and so on. It is true that the book is planned in this way, and that a great deal of its bulk is composed of selections from Rutherford's own writings. It is, however, equally true—and herein lie the merits of the book—that Professor Eve has an intimate and scholarly knowledge of Rutherford's life work, and that with him the selection and arrangement of material, and the writing of the connecting narrative, has obviously been a labour of love.

Inevitably the biography will convey most to those who knew Rutherford best—and it contains a great deal of matter which will be new even to many of his intimates. To those who had not the privilege of knowing Rutherford in the flesh it will at least give some impression of many sides of a most remarkable personality, as well as of the literally amazing range and importance of his scientific achievement.

In a brief Foreword, which is typically graceful and sincere, Lord Baldwin pays a fitting tribute to Rutherford's character and influence. H. R. R.

Introduction to Atomic Physics. By H. Semat. [Pp. xv+360.] (Farran & Rinehart, Inc., New York, 1939. Price $3.50.)

This book claims to be no more than the expanded notes of a one-semester lecture course (three hours per week) delivered to second-year students of physics. It deals briefly, in about 350 pages of fairly large type, with most aspects of atomic physics and with the elements of electromagnetism. Like so many lecture courses (including the reviewer's own) and most students' text-books on atomic physics, it has the defect, in the theoretical portions, of taking the student up to the point at which the reasoning becomes difficult, and leaving him there. It could hardly be otherwise, given the limitations of time and space imposed on such courses and books, and the capacity of the average student.

There is nothing very profound in this book, but it is a good specimen of its type. The topics are well selected, and handled in the manner of an experienced teacher who is interested in his job. The book is unpretentious and pleasing, and wholesome discipline is provided for the reader in the form of a good selection of numerical exercises.

The Mathematical Theory of Non-Uniform Gases. By S. CHAPMAN and T. G. GOWLING. [Pp. xxiii+404.] (Cambridge University Press. Price 30*s.*)

THIS very severe book deals entirely with the conclusions which can be reached by the mathematical methods of the kinetic theory regarding the properties of gases when either the velocity, the pressure, or the temperature is variable from place to place. This problem, the problem of transport phenomena, is among the most difficult of those to which the kinetic theory can be applied, and it is most useful to have for the first time a collected account of the modern work and results—largely due to the senior author—in this field. In the main the method is that of the velocity distribution function, the more elementary free-path methods having been found inadequate for these problems. The mathematical symbolism used by the authors is a logical one, but by no means easy to remember, nor can it be simple for the printer or the proof corrector, involving as it does the use of multiple symbols placed over a letter and the distinguishing of several varieties of type. Their work, however, seems to have been done perfectly. The different phenomena—viscosity, diffusion, and conductivity—are taken in turn, thermal diffusion appearing naturally as a term in the mass-transport equation, just as does the more familiar diffusion due to differences of concentration.

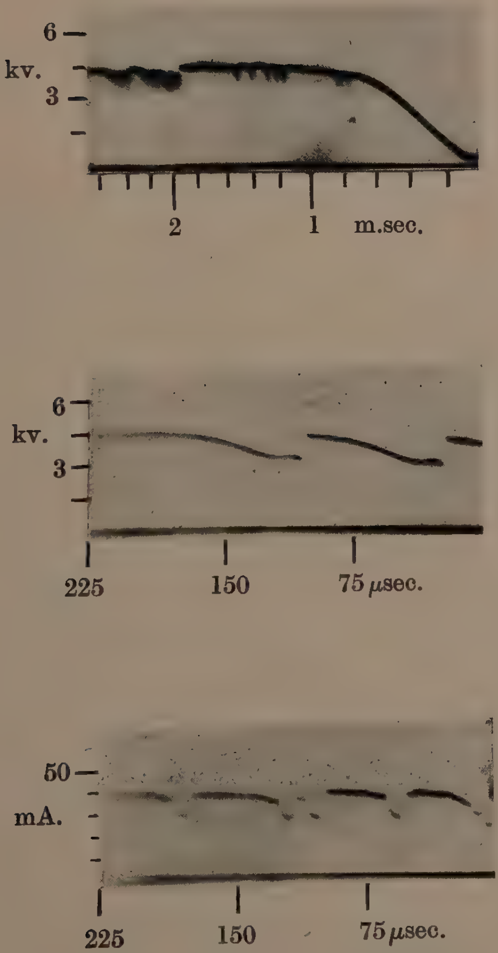
After deducing as much as possible from the general distribution function, different atomic models are taken in turn, including the rough sphere, which permits representation of the exchange of translatory with other forms of energy. Very useful chapters for the experimentalist are interpolated, where the results are collected and compared with experiment. If criticism is not ungracious where so much of value is provided, it might be directed to the undiscriminating choice of experimental results; even Planck's constant is stated to four figures as $6{\cdot}554\times10^{-27}$, although this must be nearly 1 per cent. low.

The later chapters of the book deal with special branches of the subject. Dense gases are treated with some thoroughness in chapter 16, and the modifications to classical theory necessitated by quantum theory occupy the next chapter. This is perhaps less detailed and full than one could have hoped. It contains, for example, an account of the earlier theory of electrons in metals, but abandons it at the point of most interest, the introduction of the Bloch-Brillouin theory. Finally comes a chapter on electromagnetic phenomena in ionized gases. The applications of this theory to astronomical phenomena have been largely due to Professor Chapman himself, and it is not surprising that the chapter is a very interesting one. The book concludes with appendices, of which the one giving a historical survey of the subject is of great value and would interest many to whom the mathematical difficulties of the book itself would present great obstacles.

J. H. AWBERY.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

Fig. 3.



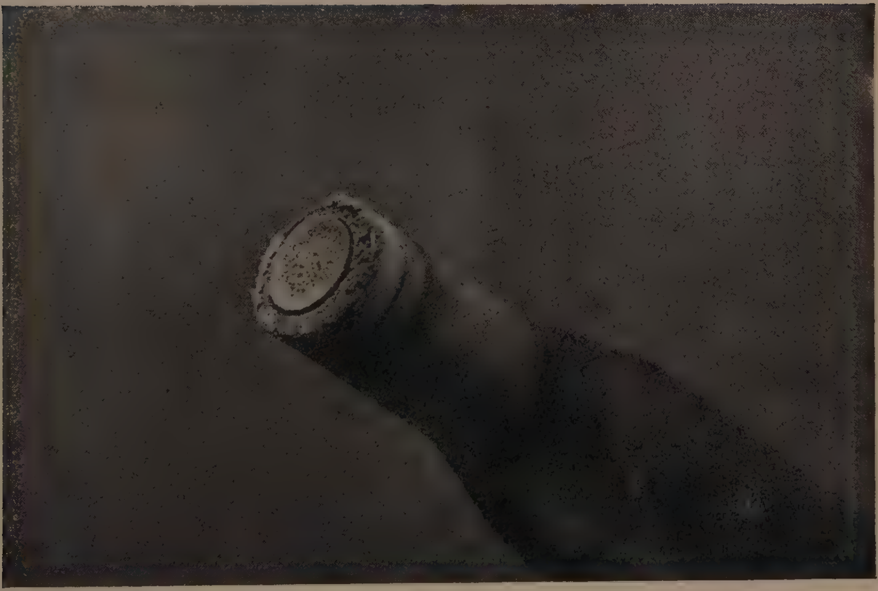
(*a*) Voltage.

(*b*) Voltage.

(*c*) Current.

Cathode-ray oscillograms.

Fig. 6.



Cathode after several hundred discharges.

XIV. *On the Mutual Electrical Energy of Two Colloidal Particles : General Theory, using Approximate Debye-Hückel Equation.*

By S. Levine †, Department of Colloid Science, Cambridge University, and G. P. Dube ‡, Institut Henri Poincaré, Paris §.

[Received August 15, 1939.]

Introduction.

In two recent papers ||, one of us developed a method of calculating the mutual electrical energy of two (charged) colloidal particles immersed in an aqueous electrolytic solution, making use of the approximate Debye-Hückel equation in the theory of electrolytes. Some applications of the results were given, and the importance of such calculations in determining the stability properties of colloidal solutions was emphasized. Only metallic particles were considered and a first and second approximation obtained. Since the method previously adopted to calculate the energy proves to be rather laborious, it is found desirable to develop an alternative method which permits the determination of the higher approximations without too much difficulty. In addition, the case of insulating particles will be treated.

The following problem is to be solved. We consider two identical spherical colloidal particles O_1 and O_2 each of radius a at a distance R apart (fig. 1) immersed in water containing an electrolyte. The potential ψ at any point in the water is assumed to satisfy the approximate Debye-Hückel equation

$$\nabla\psi=\kappa^2\psi, \quad \ldots \ldots \quad (1)$$

where κ is the characteristic quantity in the Debye-Hückel theory, being proportional to the square root of the electrolyte concentration. The function ψ is well-behaved, symmetrical with respect to O_1 and O_2, and vanishes at infinity. It is required to solve (1) and then to evaluate the following expression

$$F_2=-\frac{D\kappa^2}{8\pi}\int_V \psi^2 dv+Q\overline{\Psi}, \quad \ldots \ldots \quad (2)$$

---

† Present address, Dept. of Physics, University of Toronto, Canada.

‡ Present address, Rajput College, Agra, India.

§ Communicated by Prof. R. H. Fowler, F.R.S.

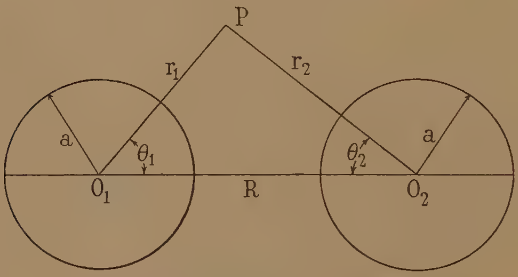
|| S. Levine, Proc. Roy. Soc. Lond. A, 170, 145, 165 (1939).

representing the electrical energy associated with the two particles. Here the integral is taken over the region of space V outside the two particles, D is the dielectric constant of the water medium, Q the charge on each particle, and $\overline{\Psi}$ is the average value of ψ over the surface of each particle. Two sets of boundary conditions at the surfaces of the particles corresponding to the metallic and insulating cases are considered, and these will be specified in a later section. However, it is convenient to introduce here one condition which will hold for both cases, namely, Gauss's theorem, which reads

$$-\frac{2Q}{Da^2}=\int_{-1}^{+1}\left(\frac{\partial\psi}{\partial r_1}\right)_{r_1=a}d\mu_1=\int_{-1}^{+1}\left(\frac{\partial\psi}{\partial r_2}\right)_{r_2=a}d\mu_2, \quad . \quad . \quad . \quad (3)$$

where r_1 and r_2 are the distances from any point P in the space to O_1 and O_2 respectively, and $\mu_1=\cos\theta_1$ and $\mu_2=\cos\theta_2$, θ_1 and θ_2 being the angles

Fig. 1.



Two colloidal particles at distance R.

$\widehat{PO_1O_2}$ and $\widehat{PO_2O_1}$ respectively. We shall first evaluate (2) and then apply the boundary conditions.

In a number of papers appearing elsewhere † a detailed discussion of the physical nature of this problem is given, and numerical examples are computed. In addition, further applications with regard to stability in colloids are considered. This paper will only deal with the mathematical derivation of the mutual energy of the two particles.

1. *General Solution of Equation* (1).

The general solution of (1) may be written as

$$\left.\begin{aligned}\psi&=\sum_{n=0}^{\infty}A_n^*\phi_n^*,\\ \phi_n^*&=P_n(\mu_1)K_{n+\frac{1}{2}}(\kappa r_1)r_1^{-\frac{1}{2}}+P_n(\mu_2)K_{n+\frac{1}{2}}(\kappa r_2)r_2^{-\frac{1}{2}},\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (4)$$

---

† S. Levine and G. P. Dube, Trans. Far. Soc. xxxv. pp. 1125, 1141 (1939); also forthcoming issue on "The Electric Double Layer." Also G. P. Dube and S. Levine, *Comptes Rendus*, ccviii. p. 1812 (1939) (preliminary note).

where $P_n(\mu)$ is the ordinary Legendre polynomial of integral order n and $K_\nu(z)$ is the modified Bessel function of the second kind †

$$K_\nu(z)=\tfrac{1}{2}\pi i^{\nu+1}H_\nu^{(1)}(iz)=\frac{\pi}{2}\cdot\frac{(i^\nu J_{-\nu}(iz)-i^{-\nu}J_\nu(iz))}{\sin\nu\pi}, \quad . \quad . \quad . \quad (5)$$

where $J_\nu(z)$ is the ordinary Bessel function of the first kind. We find it convenient to introduce

$$u_1=\kappa r_1,\quad u_2=\kappa r_2,\quad \tau=\kappa a,\quad x=\kappa R,\quad s=R/a, \quad . \quad . \quad . \quad (6)$$

these all being dimensionless quantities. Writing $A_n=\kappa^{\frac{1}{2}}A_n^*$ and $\phi_n=\kappa^{-\frac{1}{2}}\phi_n^*$, (4) becomes

$$\left.\begin{aligned}\psi&=\sum_{n=0}^{n}A_n\phi_n,\\ \phi_n&=P_n(\mu_1)K_{n+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}+P_n(\mu_2)K_{n+\frac{1}{2}}(u_2)u_2^{-\frac{1}{2}}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (7)$$

Another form for (7) is obtained by means of the relation

$$P_n(\mu_2)K_{n+\frac{1}{2}}(u_2)u_2^{-\frac{1}{2}}=\sum_{r=0}^{\infty}C_{nr}P_r(\mu_1)K_{r+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}+\sum_{r=0}^{\infty}D_{nr}P_r(\mu_1)I_{r+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}, \quad . \quad . \quad . \quad (8)$$

where $I_\nu(z)$ is the Bessel function of purely imaginary argument

$$I_\nu(z)=i^{-\nu}J_\nu(iz). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The equation (8) is true because both sides are solutions of (1), the series on the right being the general form of the solution which is valid in any given finite region of space. We proceed to find the coefficients C_{nr} and D_{nr} by making use of an expansion due to Gegenbauer.

It is necessary to distinguish two cases, namely, $u_1<x (r_1<R)$ and $u_1>x (r_1>R)$. If $u_1<x$ then ‡

$$\frac{K_{n+\frac{1}{2}}(u_2)}{u_2^{n+\frac{1}{2}}}=\frac{2^{n+\frac{1}{2}}\Gamma(n+\frac{1}{2})}{(xu_1)^{n+\frac{1}{2}}}\sum_{s=0}^{\infty}(n+s+\tfrac{1}{2})K_{n+s+\frac{1}{2}}(x)I_{n+s+\frac{1}{2}}(u_1)C_s^{n+\frac{1}{2}}(\mu_1), \quad (10)$$

where $C_s^\nu(\mu_1)$ is given by

$$(1-2\alpha t+\alpha^2)^{-\nu}=\sum_{s=0}^{\infty}C_s^\nu(t)\alpha^s,\quad C_s^{n+\frac{1}{2}}(1)=\frac{(2n+s)\,!}{(2n)\,!\,s\,!}. \quad . \quad . \quad (11)$$

---

† Throughout this paper, we shall use the notation of Watson, 'Theory of Bessel Functions.' See pp. 77–80, 132–140, 358–365.

‡ In the notation of Watson,

$$Z=\kappa R,\quad z=\kappa r,\quad \varpi=\kappa r_2=(Z^2+z^2-2Zz\cos\phi)^{\frac{1}{2}},\quad \phi=\theta,\quad |z|<|Z|$$

$$\frac{K_\nu(\varpi)}{\varpi^\nu}=2^\nu\Gamma(\nu)\sum_{m=0}^{\infty}\frac{(\nu+m)K_{\nu+m}(Z)I_{\nu+m}(z)}{Z^\nu z^\nu}C_m^\nu(\cos\phi).$$

If we now put $\mu_1=\mu_2=1$, *i. e.*, $r_2=R-r_1(u_2=x-u_1)$ and expand the binomial $u_2^n=(x-u_1)^n$, then it is readily verified that (10) may be written as

$$\frac{K_{n+\frac{1}{2}}(x-u_1)}{(x-u_1)^{\frac{1}{2}}}=\frac{2^{n+\frac{1}{2}}\Gamma(n+\frac{1}{2})}{(2n)!\,x^{\frac{1}{2}}}\sum_{p=0}^{n}(-1)^p\binom{n}{p}x^{-p}\sum_{s=0}^{\infty}\frac{(2n+s)!}{s!}$$
$$\times\frac{(n+s+\frac{1}{2})K_{n+s+\frac{1}{2}}(x)I_{n+s+\frac{1}{2}}(u_1)}{u_1^{n+\frac{1}{2}-p}}. \quad (12)$$

By a suitable transformation of the recurrence relation quoted in the footnote † and taken from Watson, we can show that

$$\frac{I_{n+s+\frac{1}{2}}(u_1)}{u_1^{n-p}}=\frac{(n-p)!}{2^{n-p}}\sum_{t=0}^{n-p}\frac{(-1)^t(s+p+2t+\frac{1}{2})}{t!\,(n-p-t)!}\,\frac{\Gamma(s+p+t+\frac{1}{2})}{\Gamma(n+s+t+\frac{3}{2})}I_{s+p+2t+\frac{1}{2}}(u_1). \quad . \quad . \quad . \quad (13)$$

On substituting (13) into (11) and interchanging twice the orders of summation, we obtain

$$\frac{K_{n+\frac{1}{2}}(x-u_1)}{(x-u_1)^{\frac{1}{2}}}=\sum_{s=0}^{\infty}\sum_{p=0}^{n}\sum_{t=0}^{n-p}B_{n,s\,p,t}I_{s+p+2t+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}, \quad . \quad . \quad (14)$$

$$=\sum_{r=0}^{\infty}I_{r+2}(u_1)u_1^{-\frac{1}{2}}\left(\sum_{s=\alpha}^{r}\sum_{p=\beta}^{\gamma}B_{n,s,p,\frac{1}{2}(r-p-s)}\right); \quad . \quad (15)$$

where

$$B_{n,s,p,t}=(-1)^{p+t}\left(\frac{2}{x}\right)^{p+\frac{1}{2}}\binom{2n+s}{2n}\binom{n}{p}\binom{n-p}{t}(n+s+\tfrac{1}{2})(s+p+2t+\tfrac{1}{2})$$
$$\times\frac{\Gamma(n+\frac{1}{2})\Gamma(s+p+t+\frac{1}{2})}{\Gamma(n+s+t+\frac{3}{2})}K_{n+s+\frac{1}{2}}(x). \quad (16)$$

In (15) $r=p+s+2t$, $\beta=1$ when $r-s$ is odd, $\beta=0$ when $r-s$ is even, $\gamma=r-s$ for $r\leqslant n+s$, $\gamma=2n+r-s$ for $r\geqslant n+s$, $\alpha=0$ for $r\leqslant 2n$ and $\alpha=r-2n$ for $r\geqslant 2n$.

On the other hand, putting $\mu_1=\mu_2=1$ (8) becomes

$$\frac{K_{n+\frac{1}{2}}(x-u_1)}{(x-u_1)^{\frac{1}{2}}}=\sum_{r=0}^{\infty}C_{nr}K_{r+2}(u_1)u_1^{-\frac{1}{2}}+\sum_{r=0}^{\infty}D_{nr}I_{r+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}. \quad . \quad (17)$$

Since this expansion is unique, it must be identical with (15), *i. e.*, when $u_1<x$,

$$\left.\begin{aligned}C_{nr}&=0,\\ D_{nr}&=\sum_{s=\alpha}^{r}\sum_{p=\beta}^{\gamma}B_{n,s,p,\frac{1}{2}(r-p-s)}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

---

† According to Watson,

$$\frac{J_{\mu+m}(z)}{z^m}=\frac{m!}{2^m}\sum_{n=0}^{m}\frac{\mu+2n}{n!\,(m-n)!}\,\frac{\Gamma(\mu+n)}{\Gamma(\mu+m+n+1)}J_{\mu+2n}(z),$$

from which (13) is obtained by substituting (9) and putting $m=n-p$, $\mu+m=n+s+\frac{1}{2}$ and hence $\mu=p+s+\frac{1}{2}$. In a similar fashion equation (20) follows on substituting (5).

For the case $u_1>x$, we must use instead of (10)

$$\frac{K_{n+\frac{1}{2}}(u_2)}{u_2^{n+\frac{1}{2}}}=\frac{2^{n+\frac{1}{2}}\Gamma(n+\frac{1}{2})}{(u_1x)^{n+\frac{1}{2}}}\sum_{s=0}^{\infty}(n+s+\tfrac{1}{2})I_{n+s+\frac{1}{2}}(x)K_{n+s+\frac{1}{2}}(u_1). \quad . \quad (19)$$

This follows from the formula in the footnote † on p. 107 if we interchange z and Z, which is permissible, since ϖ is symmetrical with respect to z and Z. The procedure is now quite similar to that adopted for $u_1<x$. Firstly, we make use of the formula (similar to (13))

$$(-1)^{n-p}\frac{K_{n+s+\frac{1}{2}}(u_1)}{u_1^{n-p}}$$

$$=\frac{(n-p)!}{2^{n-p}}\sum_{t=0}^{n-p}\frac{(-1)^t}{t!}\frac{(s+p+2t+\frac{1}{2})}{(n-p-t)!}\frac{\Gamma(s+p+t+\frac{1}{2})}{\Gamma(n+s+t+\frac{3}{2})}K_{s+p+2t+\frac{1}{2}}(u_1). \qquad (20)$$

On putting $\mu_1=-\mu_2=1$, *i. e.*, $r_2=r_1-R$ $(u_2=u_1-x)$, substituting (20) into (19), and interchanging twice the summation as in (15), we obtain an expression for (19) which, on comparison with the identical series resulting from (8) when $\mu_1=-\mu_2=1$, yields

$$\frac{(-1)^nK_{n+\frac{1}{2}}(u_1-x)}{(u_1-x)^{\frac{1}{2}}}=\sum_{r=0}^{\infty}C'_{nr}K_{r+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}+\sum_{r=0}^{\infty}D'_{nr}I_{r+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}, \qquad (21)$$

where

$$\left.\begin{aligned}D'_{nr}&=O,\\ C'_{nr}&=\sum_{s=\alpha}^{r}\sum_{p=\beta}^{\gamma}C_{n,\,s,\,p,\,\frac{1}{2}(r-p-s)},\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

with

$$C_{n,\,s,\,p,\,t}=(-1)^p\frac{I_{n+s+\frac{1}{2}}(x)}{K_{n+s+\frac{1}{2}}(x)}B_{n,\,s,\,p,\,t}. \quad . \quad . \quad . \quad . \quad . \quad (23)$$

We can now write ψ in a form different from (7). It has just been shown that

$$P_n(\mu_2)K_{n+\frac{1}{2}}(u_2)u_2^{-\frac{1}{2}}=\sum_{r=0}^{\infty}D_{nr}P_r(\mu_1)I_{r+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}, \quad r_1<R, \quad . \quad (24\text{ i})$$

$$=\sum_{r=0}^{\infty}C'_{nr}P_r(\mu_1)K_{r+\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}, \quad r_1>R. \quad . \quad (24\text{ ii})$$

Substituting (24) in (7) and interchanging the order of summation we have

$$\psi=\sum_{r=0}^{\infty}P_r(\mu_1)u_1^{-\frac{1}{2}}\{A_rK_{r+\frac{1}{2}}(u_1)+I_{r+\frac{1}{2}}(u_1)\sum_{n=0}^{\infty}A_nD_{nr}\}, \quad r_1<R, \quad . \quad (25\text{ i})$$

$$=\sum_{r=0}^{\infty}P_r(\mu_1)u_1^{-\frac{1}{2}}\{A_rK_{r+\frac{1}{2}}(u_1)+K_{r+\frac{1}{2}}(u_1)\sum_{n=0}^{\infty}A_nC'_{nr}\}. \quad r_1>R. \quad . \quad (25\text{ ii})$$

It will be found that in evaluating the integral in (2), both forms (7) and (25) need to be used.

2. *Evaluation of F_2.*

Instead of substituting the infinite series (7) in (2) it is convenient to consider the sum of the first N+1 terms

$$\psi_{N+1}=\sum_{n=0}^{N} A_n\phi_n=A_0\Big(\phi_0+\sum_{n=1}^{N}\lambda_n\phi_n\Big), \quad . \quad . \quad . \quad . \quad . \quad (26)$$

introducing

$$\lambda_n=A_n/A_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

In place of the integral in (2) we now wish to evaluate

$$\int_V \psi^2_{N+1}dv=\frac{2\pi^2 A_0^2}{\kappa^3}e^{-2\tau}\Big(\delta_{00}+2\sum_{n=1}^{N}\lambda_n\delta_{0n}+\sum_{n=1}^{N}\lambda_n^2\delta_{nn}+2\sum_{n<m=1}^{N}\lambda_n\lambda_m\delta_{nm}\Big), \quad (28)$$

where

$$\delta_{nm}=\frac{\kappa^3}{2\pi^2}e^{2\tau}\int_V \phi_n\phi_m dv. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

We proceed to develop suitable expressions for δ_{nm}.

The case $n\neq m$ will be treated first. It is found convenient to express †

$$\delta_{nm}=\delta'_{nm}+\delta''_{nm}+\delta'''_{nm}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where

$$\delta'_{nm}+\delta''_{nm}=\frac{\kappa^3}{\pi^2}e^{2\tau}\int_V P_n(\mu_1)P_m(\mu_1)K_{n+\frac{1}{2}}(u_1)K_{m+\frac{1}{2}}(u_1)u_1^{-1}\,dv \quad . \quad (31)$$

and

$$\delta'''_{nm}=\frac{\kappa^3}{2\pi^2}e^{2\tau}\int_V \{P_n(\mu_1)P_m(\mu_2)K_{n+\frac{1}{2}}(u_1)K_{m+\frac{1}{2}}(u_2)$$
$$+P_n(\mu_2)P_m(\mu_1)K_{n+\frac{1}{2}}(u_2)K_{m+\frac{1}{2}}(u_1)\}u_1^{-\frac{1}{2}}u_2^{-\frac{1}{2}}\,dv, \quad . \quad (32)$$

substituting (7) for ϕ_n. The two terms in the integrand in (32) are equal, but we require to keep them separate so as to retain the symmetry.

To evaluate (31) polar coordinates r_1, θ_1, ϕ_1 are introduced, ϕ_1 being the azimuthal angle about the line O_1O_2, so that

$$dv=r_1^2\sin\theta_1\,d\theta_1\,d\phi_1\,dr_1=r_1^2\,d\mu_1\,d\phi_1\,dr_1=\frac{1}{\kappa^3}u_1^2\,d\mu_1\,d\phi_1\,du_1.$$

The limits of integration are given by

$$\int_0^{2\pi}d\phi_1\Big(\int_0^{\infty}dr_1\int_0^{\pi}d\theta_1-\int_{R-a}^{R+a}dr_1\int_0^{\cos\overline{\theta_1}}d\theta_1\Big)$$
$$=\int_0^{2\pi}d\phi_1\Big(\int_{\tau}^{\infty}du_1\int_{-1}^{+1}d\mu_1-\int_{x-\tau}^{x+\tau}du_1\int_{\overline{\mu}_1}^{1}d\mu_1\Big), \quad . \quad (33)$$

where $\cos\overline{\theta_1}=(R^2+r_1^2-a^2)/2Rr_1=\mu_1=(x^2+u_1^2-\tau^2)/2xu_1.$

---

† In the earlier papers (footnote ‖ p. 105) we used the notation $\delta^{(')}_{nm}$, $\delta^{('')}_{nm}$, $\delta^{(''')}_{nm}$, (and also $\delta^{(')}_{\alpha n}$).

The first integral in (33) is taken over all space surrounding the particle O_1. We must subtract from this the integral taken inside the particle O_1, which is given by the second term in (33). Introducing the set of limits (33) into (31) and making use of the orthogonality (and normalization) conditions

$$\int_{-1}^{+1} P_n(\mu_1)P_m(\mu_1)d\mu_1 = \left.\begin{matrix} 0, & n\neq m, \\ 2/(2n+1), & n=m, \end{matrix}\right\} \quad . \quad . \quad . \quad . \quad (34)$$

it is readily seen that the first set of limits gives zero contribution. The second set (integral inside O_1) leads to the expression

$$-\frac{2}{\pi}e^{2\tau}\sum_{\alpha=p+1}^{n+m+1}\frac{c^{nm}_{\alpha-1}}{\alpha}\int_{x-\tau}^{x+\tau} u_1 K_{n+\frac{1}{2}}(u_1)K_{m+\frac{1}{2}}(u_1)(1-\bar{\mu}_1^{\alpha})\,du_1. \quad . \quad (35)$$

To obtain the form (35), we make use of the fact that $P_n(\mu_1)$ is a polynomial in μ_1 of degree t where $t=n/2$ or $(n-1)/2$, whichever is an integer so that we may write

$$P_n(\mu_1)P_m(\mu_1) = \sum_{k=0}^{v} c_{n+m-2k}\mu_1^{n+m-2k} = \sum_{\alpha=p+1}^{n+m+1} c^{nm}_{\alpha-1}\mu_1^{\alpha-1}, \quad . \quad . \quad . \quad (36)$$

where

$$\alpha=n+m+1-2k(\alpha=p+1, \quad p+2\ldots, n+m+1), \quad v=(n+m)/2,$$

and $p=0$ when n and m are both even, $v=(n+m-1)/2$ and $p=1$ when n is odd and m even, and $v=(n+m-2)/2$ and $p=2$ when n and m are both odd. We have substituted (36) into (31) and then integrated with respect to μ_1 from $\mu_1=\bar{\mu}_1$ to $\mu_1=1$.

The first term in (35) reads

$$-\frac{2e^{2\tau}}{\pi}S_{nm}\{I^{nm}_1(x+\tau)-I^{nm}_1(x-\tau)\}, \quad S_{nm}=\sum_{\alpha=p+1}^{n+m+1}\frac{c^{nm}_{\alpha-1}}{\alpha}, \quad . \quad . \quad . \quad (37)$$

where $I^{nm}_\beta(z)$ is the indefinite integral

$$I^{nm}_\beta(z)=\int^{z} u^{\beta}K_{n+\frac{1}{2}}(u)K_{m+\frac{1}{2}}(u)\,du. \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

The second term in (35) becomes

$$\frac{2e^{2\tau}}{\pi}\sum_{\alpha=p+1}^{n+m+1}\frac{c^{nm}_{\alpha-1}}{\alpha}\frac{1}{(2x)^{\alpha}}\sum_{\nu=0}^{\alpha}\binom{\alpha}{\nu}(x^2-\tau^2)^{\alpha-\nu}\int_{x-\tau}^{x+\tau}u_1^{2\nu+1-\alpha}K_{n+\frac{1}{2}}(u_1)K_{m+\frac{1}{2}}(u_1)\,du, \quad . \quad . \quad . \quad (39\text{ i})$$

$$=\frac{2e^{2\tau}}{\pi}\sum_{\beta=-(n+m)}^{n+m+2}E^{nm}_{\beta}\{I^{nm}_{\beta}(x+\tau)-I^{nm}_{\beta}(x-\tau)\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39\text{ ii})$$

where (putting $x=s\tau$)

$$E^{nm}_{\beta}=\frac{1}{\tau^{\beta-1}}\sum_{\alpha=|\beta-1|}^{n+m+1}\frac{c^{nm}_{\alpha-1}}{\alpha}\frac{1}{(2s)^{\alpha}}\binom{\alpha}{\frac{1}{2}(\beta+\alpha-1)}(s^2-1)^{\frac{1}{2}(\alpha-\beta+1)}, \quad . \quad (40)$$

except when $p=1$ (n odd, m even) and $\beta=1$ and $p=2$ (n and m odd) and $\beta=0$ and 2 when the lower limit in the sum in (40) is not $|\beta-1|$

but $p+1$. In (39 i) we have expanded $\bar{\mu}_1$ as a binomial ; in (39 ii) we have put $\beta=2\nu+1-\alpha$, β taking the values $-(n+m)$, $-(n+m)+2$, $n+m-1$, $n+m+1$ and then interchanged the order of summation. In (40) α takes the values $|\beta-1|$, $|\beta-1|+2$, $n+m-1$, $n+m+1$.

We now turn our attention to (32). Since the integrand in (32) is symmetrical with respect to O_1 and O_2, it is possible to write the limits of integration in the following two forms :

$$\int_0^{2\pi} d\phi_1\Big(\int_0^{\infty} du_1\int_{-1}^{+1} d\mu_1-2\int_0^{\tau} du_1\int_{-1}^{+} d\mu_1\Big) \quad . \quad . \quad . \quad . \quad \text{(41 i)}$$

$$=\lim_{\varepsilon\to 0} 2\pi\Big(\int_0^{x-\varepsilon} du_1\int_{=1}^{+1} d\mu_1-2\int_0^{\tau} du_1\int_{-1}^{+1} d\mu_1+\int_{x+\varepsilon}^{\infty} du_1\int_{-1}^{+1} d\mu_1\Big). \quad \text{(41 ii)}$$

In (41 i) the first set of limits represents the integral over all space. The second set represents the integral inside the two particles, which is equal to twice the integral inside particle O_1. In (41 ii) the integration with respect to ϕ_1 has been carried out. To evaluate (32) it is very convenient to use (41 ii). For the first two sets of limits in (41 ii) we shall introduce the expansion (24 i) into the integrand in (32) and for the last set of limits the expansion (24 ii). Remembering the conditions (34) it is not difficult to verify that the method of integrating (32) just described finally yields

$$\begin{aligned}\delta'''_{nm}=\frac{e^{2\tau}}{\pi}\Bigg[\frac{2}{2n+1}\Bigg\{&D_{mn}\Big(\Big(\int_0^{x} du_1-2\int_0^{\tau} du_1\Big)K_{n+\frac{1}{2}}(u_1)I_{n+\frac{1}{2}}(u_1)u_1\Big)\\ &+C'_{mn}\int_x^{\infty} K^2_{n+\frac{1}{2}}(u_1)u_1\,du_1\Bigg\}\\ +\frac{2}{2m+1}\Bigg\{&D_{nm}\Big(\Big(\int_0^{x} du_1-2\int_0^{\tau} du_1\Big)K_{m+\frac{1}{2}}(u_1)I_{m+\frac{1}{2}}(u_1)u_1\Big)\\ &+C'_{nm}\int_x^{\infty} K^2_{m+\frac{1}{2}}(u_1)u_1\,du_1\Bigg\}\Bigg]. \quad (42)\end{aligned}$$

In obtaining this expression the two terms constituting the integrand in (32) have been treated separately although, as noted above, their integrals are equal. Otherwise the symmetry would have been lost and the convenient form (41 iii) for the limits of integration could not have been used. The introduction of the arbitrarily small number ϵ is necessary because the series (24 i) and (24 ii) only converge uniformly in the ranges $(0, x-\epsilon)$ and $(x+\epsilon, \infty)$ respectively, and hence can be integrated term by term. Furthermore, the great advantage in using the expansion (24), the derivation of which occupied practically the whole of the previous section, in obtaining δ'''_{nm} should be noticed. In the earlier papers we used elementary methods of calcuating δ'''_{nm}, which indeed were found to be extremely tedious.

The sum of the expressions (38) and (39 ii) is equal to (31). The manner in which we separate the two contributions δ'_{nm} and δ''_{nm} is as follows. It will be seen from the examples given in the appendix that it is possible to write

$$\frac{2}{\pi}e^{2\tau}\left\{-\mathrm{S}_{nm}\mathrm{I}_1^{nm}(y)+\sum_{\beta=-(n+m)}^{n+m+2}\mathrm{E}_\beta^{nm}\mathrm{I}_\beta^{nm}(y)\right\}, \quad . \quad . \quad . \quad (43\text{ i})$$

$$=e^{-2(y-\tau)}\mathrm{R}'_{nm}(y\,;\,s,\tau)+\mathrm{R}''_{nm}(s,\tau)\int^y\frac{e^{-2u}}{u}\,du, \quad . \quad . \quad . \quad (43\text{ ii})$$

where $\mathrm{R}'_{nm}(y\,;\,s,\tau)$ and $\mathrm{R}''_{nm}(s,\tau)$ denote rational algebraic expressions in y, s, and τ. If we now put the limits of integration $y=x+\tau=\tau(s+1)$ and $y=x-\tau=\tau(s-1)$ into the indefinite integrals in (43 i), it follows that it is possible to let

$$\delta'_{nm}=e^{2\tau(s-2)}\mathrm{R}'_{nm}(\tau\overline{s-1}\,;\,s,\tau)+e^{-2\tau s}\mathrm{R}'_{nm}(\tau\overline{s+1}\,;\,s,\tau) \quad . \quad . \quad (44)$$

and

$$\delta''_{nm}=\mathrm{R}''_{nm}(s,\tau)[\mathrm{E}\,(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})], \quad . \quad . \quad . \quad . \quad (45)$$

where

$$\mathrm{E}(2y)=\int_y^\infty\frac{e^{-2u}du}{u}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . (46)$$

Furthermore, introducing the indefinite integral

$$\mathrm{J}_n(y)=\int^y u\mathrm{K}_{n+\frac{1}{2}}(u)\mathrm{I}_{n+\frac{1}{2}}(u)du, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

and noting that from the definition for $\mathrm{I}_1^{nn}(y)$, as given by (38), $\mathrm{I}_1^{nn}(\infty)=0$ because of the factor e^{-u} in $\mathrm{K}_{n+\frac{1}{2}}(u)$, we may write (42) as

$$\delta'''_{nm}=\frac{2e^{2\tau}}{\pi}\left[\frac{1}{2n+1}\{\mathrm{D}_{mn}(\mathrm{J}_n(s\tau)+\mathrm{J}_n(0)-2\mathrm{J}_n(\tau))-\mathrm{C}'_{mn}\mathrm{I}_1^{nn}(s\tau)\}\right.$$
$$\left.+\frac{1}{2m+1}\{\mathrm{D}_{nm}(\mathrm{J}_m(s\tau)+\mathrm{J}_m(0)-2\mathrm{J}_m(\tau))-\mathrm{C}'_{nm}\mathrm{I}_1^{mm}(s\tau)\}\right]. \quad . \quad (48)$$

It is a very simple matter now to consider the case $n=m$. On carrying out the integration of (29) for δ_{nn} in the same fashion as for δ_{nm}, we find that δ_{nn} may be expressed as

$$\delta_{nn}=\delta^0_{nn}+\delta'_{nn}+\delta''_{nn}+\delta'''_{nn}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

where δ'_{nn}, δ''_{nn}, and δ'''_{nn} are obtained from δ'_{nm}, δ''_{nm}, and δ'''_{nm} respectively by putting $n=m$, and the new term δ^0_{nn} arises from integrating the form (31) (for $n=m$) over the first set of limits in (33), which, because of the orthogonality conditions (34) yielded zero when $n\neq m$, but for $n=m$ now yields

$$\delta^0_{nn}=\frac{8e^{2\tau}}{(2n+1)\pi}\int_\tau^\infty\mathrm{K}^2_{n+\frac{1}{2}}(u_1)u_1du_1=-\frac{8e^{2\tau}}{(2n+1)\pi}\mathrm{I}_1^{nn}(\tau). \quad . \quad . \quad (50)$$

making use of the normalization condition in (34). The above analysis thus gives us expressions for δ_{nm} and hence for the volume integral (28). The explicit forms of δ_{nm} for n, $m=0$, 1, 2, are listed in the appendix.

We now wish to evaluate Ψ in (2) or rather $\overline{\Psi}_{N+1}$ the average value of Ψ_{N+1} over the surface S_1 of the particle O_1 where we denote the value of ψ_{N+1} on S_1 by Ψ_{N+1}. Using the form for Ψ_{N+1} obtained by putting $u_1=\tau$ and $A_n=0$, $n \geqslant N+1$ in (25 i), it is readily seen that

$$\begin{aligned}\overline{\Psi}_{N+1}&=\frac{1}{2}\int_{-1}^{+1}\Psi_{N+1}d\mu_1=\tau^{-\frac{1}{2}}\{A_0K_{\frac{1}{2}}(\tau)+I_{\frac{1}{2}}(\tau)\sum_{n=0}^{n}A_nD_{n0}\}\\&=A_0\tau^{-\frac{1}{2}}K_{\frac{1}{2}}(\tau)(1+\delta_2^{(N)}), \qquad (51)\end{aligned}$$

where

$$\delta_2^{(N)}=\delta_2\left(1+\sum_{n=1}^{N}\lambda_n\frac{D_{n0}}{D_{00}}\right), \quad \delta_2=D_{00}\frac{I_{\frac{1}{2}}(\tau)}{K_{\frac{1}{2}}(\tau)}. \qquad (52)$$

It is possible to express A_0 in terms of the λ_n's by applying Gauss's theorem (3), which gives for $\psi=\psi_{N+1}$,

$$\begin{aligned}-\frac{2Q}{Da\tau}&=\int_{-1}^{+1}\left(\frac{\partial\psi_{N+1}}{\partial u_1}\right)_{u_1=\tau}d\mu_1=\left(\frac{\partial}{\partial u_1}\right)_{u_1=\tau}\int_{-1}^{+1}\psi_{N+1}d\mu_1\\&=\left(\frac{\partial}{\partial u_1}\right)_{u_1=\tau}\{2A_0K_{\frac{1}{2}}(u_1)u_1^{-\frac{1}{2}}+2I_{\frac{1}{2}}(u_1)\sum_{n=0}^{N}A_nD_{n0}\}\\&=-2\tau^{\frac{1}{2}}\left\{A_0K_{\frac{3}{2}}(\tau)-I_{\frac{3}{2}}(\tau)\sum_{n=0}^{N}A_nD_{n0}\right\}. \qquad (53)\end{aligned}$$

Here we have used the form (25 i) with $A_n=0, n \geqslant N+1$ for ψ_{N+1} and have also introduced the relations

$$\frac{d}{dz}(K_{\frac{1}{2}}(z)z^{-\frac{1}{2}})=-K_{\frac{3}{2}}(z)z^{-\frac{1}{2}}, \quad \frac{d}{dz}(I_{\frac{1}{2}}(z)z^{-\frac{1}{2}})=I_{\frac{3}{2}}(z)z^{-\frac{1}{2}}. \qquad (54)$$

Then (53) may be written as

$$A_0=\sqrt{\frac{2}{\pi}}\kappa\frac{Qe^{\tau}}{D(1+\tau)}\cdot\frac{1}{1-\delta_1^{(N)}}, \qquad (55)$$

where

$$\delta_1^{(N)}=\delta_1\left(1+\sum_{n=1}^{N}\lambda_n\frac{D_{n0}}{D_{00}}\right), \quad \delta_1=D_{00}\frac{I_{\frac{3}{2}}(\tau)}{K_{\frac{3}{2}}(\tau)}. \qquad (56)$$

Substituting (55) in (28) and (51) we are able to write down the N+1th approximation $F_2^{(N+1)}$ to the electrical energy of the two particles. It has already been explained in the earlier papers that we only require the change in energy as the particles are brought together from infinite separation. Hence we need to subtract the quantity to which F_2 reduces when $R \rightarrow \infty$, which was shown to be

$$2F_1 = \frac{Q^2}{Da(1+\tau)}\left((1-\frac{\tau}{2(1+\tau)}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (57)$$

The preceding calculations therefore yield the following expression for the change in electrical energy

$$F_2^{(N+1)} - 2F_1 = \frac{Q^2}{Da} f_{N+1}(s, \tau) = \frac{Q^2}{Da}\left(f_r^{(N+1)}(s, \tau) + f_a^{(N+1)}(s, \tau)\right), \quad . \qquad (58)$$

where

$$f_r^{(N+1)}(s, \tau) = \frac{\delta_1^N + \delta^N}{(1+\tau)(1-\delta_1^{(N)})}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (59)$$

and

$$f_a^{(N+1)}(s, \tau) = \frac{\tau}{2(1+\tau)^2}\left\{1 - \frac{\delta_{00} + 2\sum\limits_{n=1}^{N} \lambda_n \delta_{0n} + \sum\limits_{n=1}^{N} \lambda_n^2 \delta_{nn} + 2\sum\limits_{n<m=1}^{N} \lambda_n \lambda_m \delta_{nm}}{(1-\delta_1^{(N)})^2}\right\}. \quad . \quad . \quad . \qquad (60)$$

The only quantities still to be determined are the λ_n's, which will now be found by applying the boundary conditions.

3. *Evaluation of Coefficients* λ_n.

We shall first consider the case of two insulating particles the charge on the surface of each particle being evenly distributed. Then we need to solve (1) simultaneously with Laplace's equation for the electric potential χ inside the particles, namely,

$$\nabla^2 \chi = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (61)$$

The boundary conditions at the surface S_1 of particle O_1 are

$$\psi = \Psi \equiv X = \chi \quad \text{at} \quad r_1 = a, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (62)$$

$$\left(\frac{\partial \psi}{\partial \theta_1}\right)_{r_1=a} = \left(\frac{\partial \chi}{\partial \theta_1}\right)_{r_1=a}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (63)$$

$$\frac{Q}{4\pi a^2} = -\frac{D}{4\pi}\left(\frac{\partial \psi}{\partial r_1}\right)_{r_1=a} + \frac{D'}{4\pi}\left(\frac{\partial \chi}{\partial r_1}\right)_{r_1=a}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (64)$$

where D' is the dielectric constant of the particles; analogous conditions hold at the surface of particle O_2. The appropriate solution of (61), which is finite at the centres of the two spheres, is

$$\chi = \sum_{n=0}^{N} B_n^* P_n(\mu_1) r_1^n = \sum_{n=0}^{\infty} B_n P_n(\mu_1) u_1^n, \quad \text{say}, \quad r_1 \leqslant a, \quad . \quad . \qquad (65\,\text{i})$$

$$= \sum_{n=0}^{N} B_n^* P_n(\mu_2) r_2^n = \sum_{n=0}^{\infty} B_n P_n(\mu_2) u_2^n, \quad \text{say}, \quad r_2 \leqslant a. \quad . \quad . \qquad (65\,\text{ii})$$

Instead of dealing with the infinite series it is again convenient to consider the first N+1 terms which will be denoted by χ_{N+1}. On the surface S_1 of O_1, we shall use the notation $X_{N+1}=\chi_{N+1}$.

Since in actual applications we shall always work with a finite number of terms in the series for ψ and χ it is impossible to satisfy the conditions (62)–(64) exactly. These must therefore be replaced by some approximate conditions leading to reasonable values for the coefficients A_n and B_n. When using the exact solutions for ψ and χ ($N=\infty$) it is evident that the condition (63) follows from (62); in other words (62) and (64) are sufficient to determine the coefficients uniquely. This suggests that we consider (62) and (64) only when dealing with N+1 terms. Again, with the infinite series, the coefficients A_n and B_n would be found by multiplying both sides of (62) and (64) by $P_r(\mu_1)$, $r=0, 1, 2$, and integrating in both cases from $\mu_1=-1$ to $\mu_1=+1$, making use of the orthogonality conditions (34). This also suggests that we adopt the same procedure with the N+1 terms in ψ_{N+1} and χ_{N+1}.

Proceeding along these lines, we first replace (62) by the N+1 conditions

$$0=\frac{2r+1}{2}\int_{-1}^{+1} P_r(\mu_1)(\Psi_{N+1}(\mu_1)-\chi_{N+1}(\mu_1))\,d\mu_1$$

$$=A_rK_{r+\frac{1}{2}}(\tau)\tau^{-\frac{1}{2}}+\Big(\sum_{n=0}^{N} A_nD_{nr}\Big)I_{r+\frac{1}{2}}(\tau)\tau^{-\frac{1}{2}}-B_r\tau^r. \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

Just as in the evaluation of $\bar{\Psi}_{N+1}$, we have used (25 i) for Ψ_{N+1} and (65 i) for χ_{N+1}, putting $u_1=\tau$ and $A_n=0$, $n\geqslant N+1$. Introducing $\psi=\psi_{N+1}$, $\chi=\chi_{N+1}$, and $u_1=\kappa r_1$ we shall replace (64) by the N+1 conditions

$$\frac{Q}{\kappa a^2}\int_{-1}^{+1} P_r(\mu_1)\,d\mu_1=\int_{-1}^{+1}\Big\{-D\Big(\frac{\partial\psi_{N+1}}{\partial u_1}\Big)_{u_1=\tau}$$

$$+D'\Big(\frac{\partial\chi_{N+1}}{\partial u_1}\Big)_{u_1=\tau}\Big\}P_r(\mu_1)\,d\mu_1, \quad . \quad (67)$$

or

$$0=\frac{D\tau^{-\frac{1}{2}}}{2r+1}\Big[A_r\{rK_{r-\frac{1}{2}}(\tau)+(r+1)K_{r+\frac{3}{2}}(\tau)\}$$

$$+\Big(\sum_{n=0}^{N} A_nD_{nr}\Big)\{(r+1)I_{r+\frac{3}{2}}(\tau)+rI_{r-\frac{1}{2}}(\tau)\}\Big]$$

$$+D'r\tau^{r-1}B_r, \quad r=1, \ldots, \quad N. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (68)$$

When $r=0$, (68) reduces to (3) or (55) which expresses A_0 in terms of the λ_n's. When $r>0$ (68) is obtained from (67) in a similar manner to that employed in deriving (53). The order of integration and differentiation in (67) is interchanged, the expressions (25 i) and (65 i) with $A_n=0$,

$n \geqslant N+1$, are again substituted for ψ_{N+1} and χ_{N+1} respectively, and use is made of the relations

$$\left.\begin{aligned} \frac{d}{dz}(K_{r+\frac{1}{2}}(z)z^{-\frac{1}{2}}) &= -\frac{z^{-\frac{1}{2}}}{(2r+1)}(rK_{r-\frac{1}{2}}(z)+(r+1)K_{r+\frac{3}{2}}(z)), \\ \frac{d}{dz}(I_{r+\frac{1}{2}}(z)z^{-\frac{1}{2}}) &= \frac{z^{-\frac{1}{2}}}{(2r+1)}(rI_{r-\frac{1}{2}}(z)+(r+1)I_{r+\frac{3}{2}}(z)), \end{aligned}\right\} \quad . \quad (69)$$

which are readily derived by employing the ordinary recurrence relations for $I_\nu(z)$ and $K_\nu(z)$. If we introduce $\lambda_n = A_n/A_0$ and $\nu_n = B_n/A_0$ it is possible to write (66) and (68) in the compact form

$$\sum_{n=1}^{N} a_{rn}\lambda_n + \sum_{n=0}^{N} a_{r,\,N+n+1}\nu_n = -a_{r0}, \quad r=0,\ 1,\ \ldots,\quad 2N, \quad . \quad (70)$$

where

$$\left.\begin{aligned} &a_{rn} = D_{nr}I_{r+\frac{1}{2}}(\tau), \quad a_{rr} = K_{r+\frac{1}{2}}(\tau)+D_{rr}I_{r+\frac{1}{2}}(\tau), \\ &a_{r,\,N+n+1} = 0, \quad n \neq r, \quad a_{r,\,N+n+1} = -\tau^{\frac{1}{2}(2r+1)}, \end{aligned}\right\} 0 \leqslant r \leqslant N, \quad . \quad (71)$$

$$\left.\begin{aligned} a_{rn} &= -\{(r-N)I_{r-N-\frac{1}{2}}(\tau)+(r-N+1)I_{r-n+\frac{3}{2}}(\tau)\}D_{n,\,r-N}, \\ a_{r,\,r-N} &= (r-N)K_{r-N+\frac{1}{2}}(\tau)+(r-N+1)K_{r-n+\frac{3}{2}}(\tau) \\ &\quad -\{(r-N)I_{r-n-\frac{1}{2}}(\tau)+(r-N-1)I_{r-n+\frac{3}{2}}(\tau)\}D_{r-N,\,r-N}, \\ a_{r,\,r+1} &= \tau^{r-N+\frac{1}{2}}(r-N)(2r-2N+1)\epsilon, \quad a_{r,\,N+n+1} = 0, \quad r \neq N+n, \end{aligned}\right\} N < r \leqslant 2N, \quad . \quad . \quad . \quad (72)$$

putting $\epsilon = D'/D$. The solution of (70) is

$$\lambda_n = -\sum_{r=0}^{2N-1} a_{r0}A_{rn}/|a_{rn}|, \quad \nu_n = -\sum_{r=0}^{2N-1} a_{r0}A_{r,\,N+n}/|a_{rn}|. \quad . \quad . \quad (73)$$

Here $|a_{rn}|$ is a determinant of the $2N+1$th order, $r=0, \ldots, 2N$, $n=1, \ldots, 2N+1$ and A_{rn} is the co-factor of a_{rn}.

The second case to be considered is that of two metallic particles, for which the boundary conditions are

$$\psi = \chi = \text{constant at } r_1 = a \text{ (and } r_2 = a). \quad . \quad . \quad . \quad . \quad (74)$$

Here again we shall only deal with the first $N+1$ terms in ψ_{N+1} and Ψ_{N+1} and shall replace (74) by the $N+1$ equations

$$\begin{aligned} 0 &= \frac{2r+1}{2}\int_{-1}^{+1}\Psi_{N+1}(\mu_1)P_r(\mu_1)d\mu_1 \\ &= A_rK_{r+\frac{1}{2}}(\tau)\tau^{-\frac{1}{2}} + \left(\sum_{n=0}^{N} A_nD_{nr}\right)I_{r+\frac{1}{2}}(\tau)\tau^{-\frac{1}{2}}, \quad . \quad . \quad . \quad (75) \end{aligned}$$

corresponding to (66). This can be written as

$$\sum_{n=1}^{N} a_{rn}\lambda_n = -a_{r0}, \quad r=1, 2, \ldots, N, \quad . \quad . \quad . \quad . \quad . \quad (76)$$

where a_{rn} has been defined in (71), the solutions being

$$\lambda_n = -\sum_{r=1}^{N} a_{r0} A_{rn} / |a_{rn}|, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (77)$$

where the determinant $|a_{rn}|$ is of the Nth order.

The method of determining λ_n described above is the simplest of all for both cases and is most suitable when N is sufficiently large. However, it may lead to a rather slowly converging series for F_2, and it is desirable to check the rapidity of convergence by replacing the conditions (75) (or (66) and (68)) by more appropriate ones. We shall illustrate an alternative method of determining λ_n by applying it to the simpler of the two cases, namely, to metallic particles. In order to approach the condition (74) as closely as possible we should choose A_n, $n=0, 1, \ldots, N$, such that the mean square deviation of Ψ_{N+1} from its average value, namely,

$$G_{N+1} = \frac{1}{2}\int_{-1}^{+1} (\Psi_{N+1} - \overline{\Psi}_{N+1})^2 \, d\mu_1 = \frac{1}{2}\int_{-1}^{+1} \Psi^2_{N+1} \, d\mu_1 - \overline{\Psi}^2_{N+1}, \quad . \quad (78)$$

be a minimum, subject to the condition (3) or (53). Firstly, using (51) we may write

$$\overline{\Psi}^2_{N+1} = \frac{\pi}{\tau^3} e^{-2\tau} \Big(\sum_{n=0}^{N} t_n^2 A_n^2 + 2 \sum_{n<m=0}^{N} t_n t_m A_n A_m \Big), \quad . \quad . \quad . \quad (79)$$

where

$$t_0 = \frac{\tau e^{\tau}}{\pi^{\frac{1}{2}}} \{K_{\frac{1}{2}}(\tau) + D_{00} I_{\frac{1}{2}}(\tau)\}, \quad t_n = \frac{\tau e^{\tau}}{\pi^{\frac{1}{2}}} D_{n0} I_{\frac{1}{2}}(\tau), \quad n=1, \ldots, N.$$

Secondly, if we put $\phi_n = \Phi_n$ on the surface S_1 and refer to the form (26) for ψ_{N+1}, which becomes Ψ_{N+1} on S_1, then, in order to evaluate the integral in (78), we need to consider the quantity $\Phi_n \Phi_m$. If we now use the explicit expression for Φ_n obtained by putting $u_1 = \tau$ in (7) and also introduce the orthogonality conditions (34), it is readily verified that

$$\begin{aligned} p_{nm} = p_{mn} &= \frac{\tau^3}{\pi} e^{2\tau} \int_{-1}^{+1} \Phi_n \Phi_m \, d\mu_1 \\ &= \frac{2\tau^2 e^{2\tau}}{\pi} \left[\frac{D_{mn} K_{n+\frac{1}{2}}(\tau) I_{n+\frac{1}{2}}(\tau)}{2n+1} + \frac{D_{nm} K_{m+\frac{1}{2}}(\tau) I_{m+\frac{1}{2}}(\tau)}{2m+1} \right] \\ &\quad + \frac{1}{x} \int_{x-\tau}^{x+\tau} \{K_{n+\frac{1}{2}}(u_2) K_{m+\frac{1}{2}}(u_2) P_n(\mu_2) P_m(\mu_2)\}_{u_1=\tau} \, du_2. \quad . \quad . \quad (80) \end{aligned}$$

In the integral on the right-hand side of (80) we have transformed from μ_1 to u_2 by means of the relation

$$\mu_1 = \cos\theta_1 = (R^2 + a^2 - r_2^2)/2Ra = (x^2 + \tau^2 - u_2^2)/2x\tau,$$

so that $d\mu_1 = -u_2 du_2 / x\tau$. Adopting the same method as was used to calculate (31), we introduce (36) and calculate this integral to obtain a form similar to that in (39), namely,

$$\sum_{\alpha=p}^{n+m} c_\alpha^{nm} \int_{x-\tau}^{x+\tau} \mu_2^\alpha \mathrm{K}_{n+\frac{1}{2}}(u_2)\mathrm{K}_{m+\frac{1}{2}}(u_2)\,du_2 \quad . \quad . \quad . \quad . \quad (81\text{ i})$$

$$= \sum_{\beta=-(n+m)}^{n+m} \mathrm{D}_\beta^{nm}\{\mathrm{I}_\beta^{nm}(x+\tau)-\mathrm{I}_\beta^{nm}(x-\tau)\}, \quad . \quad . \quad . \quad (81\text{ ii})$$

where ($x=s\tau$)

$$\mathrm{D}_\beta^{nm}=\frac{1}{\tau^\beta}\sum_{\alpha=|\beta|}^{n+m}\frac{c_\alpha^{nm}}{(2s)^\alpha}\binom{\alpha}{\frac{1}{2}(\alpha+\beta)}(s^2-1)^{\frac{1}{2}(\alpha-\beta)}, \quad . \quad . \quad . \quad (82)$$

except when $p=2$ (n and m odd) and $\beta=0$ when the lower limit in the sum in (82) is not $\alpha=0$, but $\alpha=2$. In (81 i) we have put

$$\mu_2^\alpha=(\cos\theta_2)^\alpha=\frac{1}{(2x)^\alpha}\left(\frac{x^2-\tau^2}{u_2^2}+u_2\right)^\alpha=\frac{1}{(2x)^\alpha}\sum_{\nu=0}^{\alpha}\binom{\alpha}{\nu}(x^2-\tau^2)^{\alpha-\nu}u_2^{2\nu-\alpha},$$

and then interchanged the order of summation with $\beta=2\nu-\alpha$ taking the values $-(n+m)$, $-(n+m)+2$, . . ., $n+m-2$, $n+m$. In this way we derive an expression for p_{nm}. To obtain the corresponding form for $n=m$ we need only put $n=m$ in (80), (81), and (82) and also add on the new term $\{2\tau^2e^{2\tau}/(2n+1)\pi\}\mathrm{K}^2_{n+\frac{1}{2}}(\tau)$ which is analogous to the term $\delta_{nn}^{(0)}$ in δ_{nn}, arising from the normalization condition in (34).

Substituting into (78), (79) and the form for $\Psi_{\mathrm{N}+1}$ resulting from (26) on putting $\phi_n=\Phi_n$ and then the expressions just derived for p_{nm} and p_{nr}, we obtain

$$\mathrm{G}'_{\mathrm{N}+1}=\sum_{n=0}^{\mathrm{N}} h_{nn}^2\mathrm{A}_n^2+2\sum_{n<m=0}^{\mathrm{N}} h_{nm}\mathrm{A}_n\mathrm{A}_m=\frac{\tau^3}{\pi}e^{2\tau}\mathrm{G}_{\mathrm{N}+1}, \quad . \quad . \quad (83)$$

where $h_{nn}=p_{nn}/2-t_n^2$, $h_{nm}=h_{mn}=p_{nm}/2-t_nt_m$. The method of undetermined multipliers is now applied to find the coefficients. It is required that $\mathrm{G}'_{\mathrm{N}+1}$ be a minimum subject to the condition (55), which may be written as

$$\mathrm{H}_{\mathrm{N}+1}\equiv\sum_{n=0}^{\mathrm{N}} a_{0n}\mathrm{A}_n-b_0=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (84)$$

where $b_0=\mathrm{Q}/\mathrm{D}a\tau^{\frac{1}{2}}$ and a_{0n} has already been defined in (72). If μ is an arbitrary multiplier then $\mathrm{L}_{\mathrm{N}+1}=\mathrm{G}'_{\mathrm{N}+1}+\mu\mathrm{H}_{\mathrm{N}+1}$ will be a minimum, *i. e.*,

$$\frac{\partial\mathrm{L}_{\mathrm{N}+1}}{\partial\mu}=\mathrm{H}_{\mathrm{N}+1}=0, \quad \frac{\partial\mathrm{L}_{\mathrm{N}+1}}{\partial\mathrm{A}_r}=2\sum_{n=0}^{\mathrm{N}} g_{rn}\mathrm{A}_n+\mu a_{0r}=0, \quad r=0,\ 1,\ .\ .\ .,\ \mathrm{N}, \quad (85)$$

from which μ can be eliminated and the following N equations for λ_n are obtained :

$$\sum_{n=1}^{\mathrm{N}} a^*_{rn}\lambda_n=-a^*_{r0}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (86)$$

with $a^*_{rn}=a_{00}h_{rn}-a_{0r}h_{0n}$. The solution of (86) is

$$\lambda^*_n=-\sum_{r=1}^{N} a^*_{r0}A^*_{rn}/|a^*_{rn}|, \quad \dots \dots \dots \quad (87)$$

A^*_{rn} being the co-factor of a^*_{rn} (period).

In the appendix the explicit expressions for λ_n when N=2 and 3 are given for both insulating and metallic particles. The corresponding values of λ^*_n for the metallic case are also computed. In the paper devoted to numerical work and appearing elsewhere † it is shown that there is not an important difference between the values of λ_n and λ^*_n for the metallic particles. Hence we considered that it was not sufficiently important to treat insulating particles by the method of least square deviation.

Appendix.

In the application of the general results obtained in this report we have only made use of the explicit expressions associated with either one, two, or three terms in the series for ψ (*i. e.*, N=1, 2, or 3). From (18) and (22) we derive ($x=s\tau$)

$$D_{0r}=B_{0,\,r,\,0,\,0}=\left(\frac{2\pi}{s\tau}\right)^{\frac{1}{2}}(r+\tfrac{1}{2})K_{r+\frac{1}{2}}(s\tau),\quad D_{n0}=B_{n,\,0,\,0,\,0}=\left(\frac{\pi}{2s\tau}\right)^{\frac{1}{2}}K_{n+\frac{1}{2}}(s\tau),$$

$$C'_{0r}=C'_{0,\,r,\,00}=\left(\frac{2\pi}{s\tau}\right)^{\frac{1}{2}}(r+\tfrac{1}{2})I_{r+\frac{1}{2}}(s\tau),\quad (-1)^n C'_{n0}=C'_{n,\,0,\,0,\,0}=\left(\frac{\pi}{2s\tau}\right)^{\frac{1}{2}}I_{n+\frac{1}{2}}(s\tau).$$

Instead of quoting the general expressions for the higher coefficients, only the particular ones needed here will be given. It is convenient to put

$$\rho^{\pm}_{10}(u)=1\pm\frac{1}{u},\quad \rho^{\pm}_{11}(u)=1\pm\frac{2}{u}+\frac{2}{u^2},\quad \rho^{\pm}_{20}(u)=1\pm\frac{3}{u}+\frac{3}{u^2},$$

$$\rho^{\pm}_{12}(u)=\frac{7}{2}\pm\frac{37}{2u}+\frac{45}{u^2}\pm\frac{45}{u^3}.\quad \rho^{\pm}_{21}(u)=1\pm\frac{4}{u}+\frac{9}{u^2}\pm\frac{9}{u^3},$$

$$\rho^{\pm}_{22}(u)=1\pm\frac{6}{u}+\frac{24}{u^2}\pm\frac{54}{u^3}+\frac{54}{u^4}.\quad \rho^{\pm}_{30}(u)=1\pm\frac{6}{u}+\frac{15}{u^2}\pm\frac{15}{u^3}.$$

Then the formulæ for $K_{n+\frac{1}{2}}(u)$ and $I_{n+\frac{1}{2}}(u)$ may be written as

$$K_{n+\frac{1}{2}}(u)=K_{-(n+\frac{1}{2})}(u)=\left(\frac{\pi}{2u}\right)^{\frac{1}{2}}e^{-u}\alpha_{n+\frac{1}{2}}(u),\quad I_{\pm(n+\frac{1}{2})}(u)=\left(\frac{1}{2\pi u}\right)^{\frac{1}{2}}e^{u}\beta_{\pm(n+\frac{1}{2})}(u), \quad \dots \quad (A.\ 1)$$

where, in particular,

$$\alpha_{\frac{1}{2}}(u)=1,\quad \alpha_{\frac{3}{2}}(u)=\rho^{+}_{10}(u),\quad \alpha_{\frac{5}{2}}(u)=\rho^{+}_{20}(u),\quad \beta_{\frac{1}{2}}(u)=1-e^{-2u},$$

$$\beta_{\frac{3}{2}}(u)=\rho^{-}_{10}(u)+e^{-2u}\rho^{+}_{10}(u),\quad \beta_{\frac{5}{2}}(u)=\rho^{-}_{20}(u)-e^{-2u}\rho^{+}_{20}(u),$$

$$\beta_{\frac{7}{2}}(u)=\rho^{-}_{30}(u)+e^{-2u}\rho^{+}_{30}(u).$$

---

† S. Levine and G. P. Dube, *loc. cit.*

It follows that

$$D_{nr}=\frac{\pi}{2s\tau}e^{-s\tau}d_{nr},\quad C'_{nr}=\frac{e^{s\tau}}{2s\tau}c'_{nr},\quad .\quad .\quad .\quad .\quad (A.\ 2)$$

where

$$d_{00}=1,\quad d_{01}=3\,d_{10}=3\rho^{+}_{10}(s\tau),\quad d_{11}=3\rho^{+}_{11}(s\tau),\quad d_{02}=5\,d_{20}=5\rho^{+}_{20}(s\tau),$$

$$d_{12}=\rho^{+}_{12}(s\tau),\quad d_{21}=3\rho^{+}_{21}(s\tau),\quad d_{22}=5\rho^{+}_{22}(s\tau),\quad c'_{00}=\beta_{\frac{1}{2}}(s\tau),$$

$$c'_{01}=-3\,c'_{10}=3\beta_{\frac{3}{2}}(s\tau),\quad c'_{02}=5\,c'_{20}=5\beta_{\frac{5}{2}}(s\tau),$$

$$c'_{11}=-3(\rho^{-}_{11}(s\tau)-e^{-2s\tau}\rho^{+}_{11}(s\tau)),\quad c'_{12}=-(\rho^{-}_{12}(s\tau)+e^{-2s\tau}\rho^{+}_{12}(s\tau)),$$

$$c_{21}=3(\rho^{-}_{21}(s\tau)+e^{-2s\tau}\rho^{+}_{21}(s\tau)),\quad c'_{22}=5(\rho^{-}_{22}(s\tau)-e^{-2s\tau}\rho^{+}_{22}(s\tau)).$$

From the ordinary expressions for the Legendre polynomials we have, in (36),

$$c^{00}_{0}=1,\quad c^{01}_{0}=1,\quad c^{02}_{0}=-\tfrac{1}{2},\quad c^{02}_{2}=\tfrac{3}{2},\quad c^{11}_{2}=1,\quad c^{12}_{1}=-\tfrac{1}{2},$$

$$c^{12}_{3}=\tfrac{3}{2},\quad c^{22}_{0}=\tfrac{1}{4},\quad c^{22}_{2}=-\tfrac{3}{2},\quad c^{22}_{4}=\tfrac{9}{4},$$

from which we derive

$$S^{00}=1,\quad S_{01}=\tfrac{1}{2},\quad S_{02}=0,\quad S_{12}=-\tfrac{1}{8},\quad S_{11}=\tfrac{2}{3},\quad S_{22}=\tfrac{1}{5},$$

according to (37),

$$E^{00}_{0}=\frac{\tau(s^2-1)}{2s},\quad E^{00}_{2}=\frac{1}{2s\tau},\quad E^{00}_{-1}=\frac{1}{2}\left(\frac{\tau\overline{s^2-1}}{2s}\right)^2,\quad E^{01}_{1}=\frac{s^2-1}{2s^2},$$

$$E^{01}_{3}=\frac{1}{8(s\tau)^2},\quad E^{11}_{-2}=\frac{1}{3}\left(\frac{\tau\overline{s^2-1}}{2s}\right)^3=\frac{2}{3}E^{02}_{-2},\quad E^{11}_{0}=\frac{\tau(s^2-1)^2}{8s^3},$$

$$E^{11}_{2}=\frac{s^2-1}{4s\tau^3},\quad E^{11}_{4}=\frac{1}{24(s\tau)^3}=\frac{2}{3}E^{02}_{4},\quad E^{02}_{0}=-\frac{\tau(s^2-1)(s^2+3)}{16s^3}$$

$$E^{02}_{2}=-\frac{s^2+3}{16\tau s^3},\quad E^{12}_{-3}=\frac{3}{8}\left(\frac{\tau\overline{s^2-1}}{2s}\right)^4,\quad E^{12}_{-1}=\frac{\tau(s^2-1)^2(s^2-3)}{32s^4},$$

$$E^{12}_{1}=\frac{(s^2-1((s^2-9)}{64s^4},\quad E^{12}_{3}=\frac{s^2-3}{2\tau^2(2s)^4},\quad E^{12}_{5}=\frac{3}{8}\frac{1}{(2s\tau)^4},$$

$$E^{22}_{-4}=\frac{9}{20}\left(\frac{\tau\overline{s^2-1}}{2s}\right)^5,\quad E^{22}_{-2}=\frac{1}{4}\left(\frac{\tau\overline{s^2-1}}{2s}\right)^3\frac{(s^2-9)}{(2s)^2},$$

$$E^{22}_{0}=\frac{\tau(s^2-1)(5s^4-6s^2+9)}{64s^5}=\tau^2(s^2-1)E^{22}_{2},\quad E^{22}_{4}=\frac{1}{4\tau^3}\frac{s^2-9}{(2s)^5},$$

$$E^{22}_{6}=\frac{9}{20}\frac{1}{(2s\tau)^5},$$

according to (40), and

$$D^{00}_{0}=1,\quad D^{01}_{-1}=\frac{\tau(s^2-1)}{2s},\quad D^{01}_{1}=\frac{1}{2s\tau},\quad D^{11}_{0}=\frac{s^2-1}{2s^2},$$

$$D^{11}_{-2}=\left(\frac{\tau\overline{s^2-1}}{2s}\right)^2=\frac{2}{3}D^{02}_{-2},\quad D^{11}_{2}=\frac{1}{(2s\tau)^2}=\frac{2}{3}D^{02}_{2},\quad D^{02}_{0}=-\frac{s^2+3}{8s^2},$$

$$D^{12}_{-3}=\frac{3}{2}\left(\frac{\overline{\tau s^2-1}}{2s}\right)^3, \quad D^{12}_{-1}=\frac{\tau(s^2-1)(5s^2-9)}{16s^3}, \quad D^{12}_{1}=\frac{5s^2-9}{16s^3\tau},$$

$$D^{12}_{3}=\frac{3}{16(s\tau)^3}, \quad D^{22}_{-4}=\frac{9}{4}\left(\frac{\overline{\tau s^2-1}}{2s}\right)^4, \quad D^{22}_{-2}=\frac{3\tau^2(s^2-3)(s^2-1)^2}{16s^4},$$

$$D^{22}_{0}=\frac{11s^4-30s^2+27}{32s^4}, \quad D^{22}_{2}=\frac{3(s^2-3)}{16s^4\tau^2}, \quad D^{22}_{4}=\frac{9}{64(s\tau)^4},$$

according to (82).

In order to evaluate the integrals I^{nm}_{β} and J_n defined by (38) and (47) respectively, we make use of the following types of integral computed by Watson :

$$\int^z \mathscr{C}_{\mu}(z)\overline{\mathscr{C}}_{\mu}(z)\frac{dz}{z}, \quad \int^z z\mathscr{C}_{\mu}(z)\overline{\mathscr{C}_{\mu}}(z)\,dz, \quad \int^z z^{\mu+\nu+1}\mathscr{C}_{\mu}(z)\overline{\mathscr{C}}_{\nu}(z)\,dz,$$

$$\int^z z^{-\mu-\nu-1}\mathscr{C}_{\mu+1}(z)\overline{\mathscr{C}}_{\nu+1}(z)\,dz,$$

and the reduction formula for

$$\int^z z_{\rho-1}\mathscr{C}_{\mu+1}(z)\overline{\mathscr{C}}_{\nu+1}(z)\,dz, \quad \int^z {}^{\mu+2}\mathscr{C}^2_{\nu}(z)\,dz.$$

Here $\mathscr{C}_{\mu}(z)$ and $\overline{\mathscr{C}}_{\mu}(z)$ are any two cylinder functions (one of the three kinds of Bessel functions). Now $H^1_{\nu}(iz)$ in (5) and $J_{\nu}(iz)$ in (9) are both cylinder functions, so that we can express $K_{\nu}(z)$ and $I_{\nu}(z)$ in terms of cylinder functions. The following results have been obtained :—

$$J_0=\frac{1}{4}(2z+e^{-2z}), \quad J_1=\frac{1}{4}\left\{2z\left(1+\frac{1}{z^2}\right)-e^{-2z}\left(1+\frac{2}{z}\right)\right\},$$

$$J_2=\frac{1}{4}\left\{2z\left(1+\frac{3}{2z}-\frac{3}{z^4}\right)+e^{-2z}\left(1+\frac{6}{z}\left(1+\frac{1}{z}\right)^2\right)\right\},$$

$$I^{00}_{0}=\frac{\pi}{2}E^*(2z), \quad I^{00}_{1}=-\frac{\pi}{4}e^{-2z}, \quad I^{00}_{2}=-\frac{\pi}{8}e^{-2z}(2z+1),$$

$$I^{01}_{-1}=-\frac{\pi}{4}\frac{e^{-2z}}{z^2}, \quad I^{01}_{1}=\frac{\pi}{4}(e^{-2z}+2E^*(2z)), \quad I^{01}_{3}=-\frac{\pi}{4}e^{-2z}(z+1)^2,$$

$$I^{02}_{-2}=-\frac{\pi}{8}\frac{e^{-2z}}{z^4}(2z+3), \quad I^{02}_{0}=\frac{\pi}{2}\left\{-\frac{3}{2z^2}e^{-2z}+E^*(2z)\right\},$$

$$I^{02}_{2}=\frac{\pi}{2}\left\{-\frac{e^{-2z}}{4}(2z+7)+3E^*(2z)\right\},$$

$$I^{02}_{4}=-\frac{\pi}{16}e^{-2z}(4z^3+18z^2+30z+15),$$

$$I^{11}_{-2}=-\frac{\pi}{8}\frac{e^{-2z}}{z^4}(2z+1), \quad I^{11}_{0}=-\frac{\pi}{4}\left\{\frac{e^{-2z}}{z^2}(2z+1)+2E^*(2z)\right\},$$

$$I_1^{11}=-\frac{\pi}{4}e^{-2z}\left(1+\frac{2}{z}\right),\quad I_2^{11}=\frac{\pi}{4}\left\{-e^{-2z}\left(z+\frac{5}{2}\right)+2E^*(2z)\right\},$$

$$I_4^{11}=-\frac{\pi}{16}e^{-2z}(4z^3+14z^2+18z+9),$$

$$I_{-3}^{12}=-\frac{\pi}{4}\frac{e^{-2z}}{z^6}(z+1)^2,\quad I_{-1}^{12}=-\frac{\pi}{8}\frac{e^{-2z}}{z^4}(2z^2+6z+3),$$

$$I_1^{12}=-\pi\left\{\frac{e^{-2z}}{4z^2}(z^2+6z+3)+E^*(2z)\right\},$$

$$I_3^{12}=-\frac{\pi}{2}\left\{\frac{e^{-2z}}{4}(2z^2+10z+17)+3E^*(2z)\right\},$$

$$I_5^{12}=-\frac{\pi}{4}e^{-2z}(z^2+3z+3)^2,$$

$$I_{-4}^{22}=-\frac{\pi}{4}\frac{e^{-2z}}{z^5}\left(1+\frac{7}{2z}+\frac{9}{2z^2}+\frac{9}{4z^3}\right),\quad I_{-2}^{22}=-\frac{\pi}{4}\frac{e^{-2z}}{z^3}\left(1+\frac{9}{2z}+\frac{6}{z^2}+\frac{3}{z^3}\right),$$

$$I_0^{22}=\frac{\pi}{4}\left\{-\frac{3}{z^2}e^{-2z}\left(1+\frac{3}{z}+\frac{3}{2z^2}\right)+2E^*(2z)\right\},$$

$$I_1^{22}=-\frac{\pi}{4}e^{-2z}\left\{1+\frac{6}{z}\left(1+\frac{1}{z}\right)^2\right\},$$

$$I_2^{22}=-\frac{\pi}{4}\left\{e^{-2z}z\left(1+\frac{13}{2z}+\frac{18}{z^2}+\frac{9}{z^3}\right)+6E^*(2z)\right\},$$

$$I_4^{22}=-\frac{\pi}{4}\left\{e^{-2z}\left(z^3+\frac{15}{2}z^2+\frac{45}{2}z+\frac{117}{4}\right)-18E^*(2z)\right\},$$

$$I_6^{22}=-\frac{\pi}{4}e^{-2z}\left(z^5+\frac{17}{2}z^4+32z^3+66z^2+75z+\frac{75}{2}\right).$$

We have denoted the indefinite integral $\int(e^{-2z}/z)dz$ simply by $E^*(2z)$. The integrals I_0^{00}, I_1^{01}, I_2^{11}, I_2^{02}, I_3^{12}, and I_0^{02} have to be determined by elementary methods, and use of the first three is made in evaluating a number of the integrals above.

Substituting all these formulæ in (46) and (59) the following expressions † for the various contributions to δ_{nm} are readily verified :—

$$\delta_{00}^0=1,\quad \delta_{00}'=-\frac{e^{-2\tau(s-2)}}{85\tau}\{2s\tau+2\tau-1+e^{-4\tau}(2\tau-2\tau s+1)\},$$

$$\delta_{00}''=-\frac{\tau(s^2-1)e^{2\tau}}{2s}\{E(2\tau\overline{s-1})-E(2\tau\overline{s+1})\},\quad \delta_{00}'''=\frac{e^{-\tau(s-2)}}{8\tau}\{1+\tau\overline{s-2}-e^{2\tau}\}.$$

$$\delta_{01}'=\frac{e^{-2\tau(s-2)}}{8s^2\tau^2}\{1+2\tau s-2\tau-e^{-4\tau}(1+2\tau s+2\tau)\},$$

---

† The formulæ for δ_{00}, δ_{01}, and δ_{11} have already been obtained in one of the previous papers (footnote ‖ p. 105), the notation adopted being $\delta_{00}=1+\delta_3$, $\delta_{01}=\delta_4+\delta_5$, $\delta_{11}=\delta_6+\delta_7$, $\delta_{00}+\delta_{00}''=-\mu_1$, $\delta_{00}'''=\mu_2$.

$$\delta''_{01}=-\frac{e^{2\tau}}{2}\left(\frac{1+s^2}{s^2}\right)\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\},$$

$$\delta'''_{01}=2e^{-\tau(s-2)}\left[1-\frac{(1+s\tau)}{s^2\tau^3}(1-2\tau+2\tau^2-e^{-2\tau})\right].$$

$$\delta^0_{11}=\frac{1}{3}\left(1+\frac{2}{\tau}\right),\quad \delta''_{11}=\frac{e^{2\tau}(s^2-1)}{8s^3\tau}(1-\tau^2\overline{s^2-1})\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\}$$

$$\begin{aligned}\delta'_{11}=\frac{e^{-2\tau(s-2)}}{48}&\Bigg[\left(\frac{9}{4\tau^3}-\frac{9}{2\tau^2}-\frac{1}{\tau}-3\right)\frac{1}{s^3}+\left(\frac{9}{2\tau^2}-\frac{1}{\tau}-3\right)\frac{1}{s^2}+\left(3+\frac{14}{\tau}\right)\frac{1}{s}+3\\&+\frac{(s+1)^3-32s^3}{2s^3\tau(s-1)}-e^{-4\tau}\Bigg\{\left(\frac{9}{4\tau^3}+\frac{9}{2\tau^2}-\frac{1}{\tau}+3\right)+\left(\frac{9}{2\tau^2}-3+\frac{1}{\tau}\right)\frac{1}{s^2}\\&+\left(\frac{14}{\tau}-3\right)\frac{1}{s}+3+\frac{(s-1)^3-32s^3}{2s^3\tau(s+1)}\Bigg\}\Bigg],\end{aligned}$$

$$\delta'''_{11}=e^{-\tau(s-2)}\left[1+\frac{1}{s\tau^2}\left(1+\frac{2}{s\tau}+\frac{2}{(s\tau)^2}\right)\{3\tau-2-2\tau^2+(\tau+2)e^{-2\tau}\}\right].$$

$$\begin{aligned}\delta'_{02}=\frac{e^{-2\tau(s-2)}}{32}&\Bigg[\Bigg\{1+\left(1+\frac{1}{\tau}\right)\frac{1}{s}+3\left(1-\frac{3}{\tau}+\frac{5}{2\tau^2}\right)\frac{1}{s^2}\\&+3\left(1-\frac{2}{\tau}-\frac{5}{2\tau^2}+\frac{5}{4\tau^3}\right)\frac{1}{s^3}+\frac{3(s+1)(s^2-2s+5)}{2\tau(s-1)}\Bigg\}\\&-e^{-4\tau}\Bigg\{1+\left(\frac{1}{\tau}-1\right)\frac{1}{s}+3\left(1+\frac{3}{\tau}+\frac{5}{2\tau^2}\right)\frac{1}{s^2}\\&+3\left(-1-\frac{2}{\tau}+\frac{5}{2\tau^2}+\frac{5}{4\tau^3}\right)\frac{1}{s^3}-\frac{3(s-1)(s^2+2s+5)}{2\tau(s+1)}\Bigg\}\Bigg],\end{aligned}$$

$$\delta''_{02}=-\frac{(s^2+3)e^{2\tau}}{16s^3}\left\{\tau(s^2-1)+\frac{3}{\tau}\right\}\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\},$$

$$\begin{aligned}\delta'''_{02}=\frac{2e^{-\tau(s-2)}}{s\tau}&\Bigg[(2+s\tau)+\left(1+\frac{3}{s\tau}+\frac{3}{s^2\tau^2}\right)\\&\times\left\{3-2\tau-\frac{3}{\tau}+\frac{3}{\tau^3}-e^{-2\tau}\left(1+\frac{3(\tau+1)^2}{\tau^3}\right)\right\}\Bigg].\end{aligned}$$

$$\begin{aligned}\delta'_{12}=\frac{e^{-2\tau(s-2)}}{128}&\Bigg[\frac{3}{\tau(s-1)}\left(2+\frac{1}{\tau(s-1)}\right)\left(-\frac{11}{2}+\frac{4}{s}-\frac{9}{s^2}-\frac{4}{s^3}+\frac{13}{2s^4}\right)+\frac{19}{\tau s}\\&+\left(-\frac{37}{\tau}+\frac{79}{2\tau^2}\right)\frac{1}{s^2}+\frac{3}{\tau}\left(-1-\frac{15}{\tau}+\frac{9}{\tau^2}\right)\frac{1}{s^3}\\&+\left(\frac{21}{\tau}-\frac{57}{2\tau^2}-\frac{27}{\tau^3}+\frac{27}{2\tau^4}\right)\frac{1}{s^4}-e^{-4\tau}\Bigg\{\frac{3}{\tau(s+1)}\left(2+\frac{1}{\tau(s+1)}\right)\\&\times\left(-\frac{11}{2}-\frac{4}{s}-\frac{9}{s^2}+\frac{4}{s^3}+\frac{13}{2s^4}\right)+\frac{19}{s\tau}+\left(\frac{37}{\tau}+\frac{79}{2\tau^2}\right)\frac{1}{s^2}\\&+\frac{3}{\tau}\left(-1+\frac{15}{\tau}+\frac{9}{\tau^2}\right)\frac{1}{s^3}+\left(-\frac{21}{\tau}-\frac{57}{2\tau^2}+\frac{27}{\tau^3}+\frac{27}{2\tau^4}\right)\frac{1}{s^4}\Bigg\}\Bigg],\end{aligned}$$

$$\delta''_{12}=\frac{e^{2\tau}}{32}\left\{7+\left(10+\frac{3}{\tau^2}\right)\frac{1}{s^2}-9\left(1+\frac{1}{\tau^2}\right)\frac{1}{s^4}\right\}\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\},$$

$$\begin{aligned}\delta'''_{12}=\frac{e^{-\tau(s-2)}}{s\tau}&\left[s\tau+3\left(1+\frac{1}{s\tau}\right)+\left\{\left(3-2\tau-\frac{2}{\tau}\right)+\left(1+\frac{2}{\tau}\right)e^{-2\tau}\right\}\right.\\&\left.\times\left\{1+\frac{4}{s\tau}+\frac{9}{(s\tau)^2}+\frac{9}{(s\tau)^3}\right\}\right]+\frac{e^{-\tau(s-2)}}{10s\tau}\left[\frac{7}{2}s\tau+8\left(1+\frac{1}{s\tau}\right)\right.\\&+\left\{\left(5-2\tau-\frac{6}{\tau}+\frac{6}{\tau^3}\right)-\left(1+\frac{6}{\tau}\left(1+\frac{1}{\tau}\right)^2\right)e^{-2\tau}\right\}\\&\left.\times\left(\frac{7}{2}+\frac{37}{2s\tau}+\frac{45}{(s\tau)^2}+\frac{45}{(s\tau)^3}\right)\right].\end{aligned}$$

$$\delta^0_{22}=\frac{1}{5}\left\{1+\frac{6}{\tau}\left(1+\frac{1}{\tau}\right)^2\right\},$$

$$\begin{aligned}\delta'_{22}=-\frac{e^{-2\tau(s-2)}}{128}&\left[\left\{\frac{64}{5}\left(1+\frac{6}{\tau(s-1)}\left(1+\frac{1}{\tau(s-1)}\right)^2\right)\right.\right.\\&-\frac{9}{10}\left(\frac{s+1}{s}\right)^5\left(1+\frac{7}{2\tau(s-1)}+\frac{9}{2\tau^2(s-1)^2}+\frac{9}{4\tau^3(s-1)^3}\right)\\&-\frac{(s+1)^3(s^2-9)}{2s^5}\left(1+\frac{9}{2\tau(s-1)}+\frac{6}{\tau^2(s-1)^2}+\frac{3}{\tau^3(s-1)^3}\right)\\&+\frac{3(s+1)}{s^5(s-1)}(5s^4-6s^2+9)\left(1+\frac{3}{\tau(s-1)}+\frac{3}{2\tau^2(s-1)^2}\right)\\&+\frac{5s^4-6s^2+9)(s-1)}{s^5}\left(1+\frac{13}{\tau(s-1)}+\frac{18}{\tau^2(s-1)^2}+\frac{9}{\tau^3(s-1)^3}\right)\\&+\frac{(s^2-9)(s-1)^3}{2s^5}\left(1+\frac{15}{2\tau(s-1)}+\frac{45}{2\tau^2(s-1)^2}+\frac{117}{4\tau^3(s-1)^3}\right)\\&+\frac{9(s-1)^5}{10s^5}\left(1+\frac{17}{2\tau(s-1)}+\frac{32}{\tau^2(s-1)^2}+\frac{66}{\tau^3(s-1)^3}+\frac{75}{\tau^4(s-1)^4}\right.\\&\left.\left.+\frac{75}{2\tau^5(s-1)^5}\right)\right\}\\&\left.-e^{-4\tau}\{\text{same form with }(s-1)\text{ and }(s+1)\text{ interchanged}\}\right],\end{aligned}$$

$$\begin{aligned}\delta''_{22}=\frac{e^{2\tau}}{64s^2\tau^2}&\left\{\tau^2(5s^4-6s^2+9)(\tau^2(s^2-1)-3)+\frac{9}{2}(s^2-9)\right\}\\&\times\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\},\end{aligned}$$

$$\begin{aligned}\delta'''_{22}=e^{-\tau(s-2)}&\left[1+\frac{4}{s\tau}+\frac{6}{(s\tau)^2}\left(1+\frac{1}{s\tau}\right)+\frac{1}{s\tau}\left\{\left(5-2\tau-\frac{6}{\tau}+\frac{6}{\tau^3}\right)\right.\right.\\&\left.\left.-\left(1+\frac{6}{\tau}\left(1+\frac{1}{\tau}\right)^2\right)e^{-2\tau}\right\}\left(1+\frac{6}{s\tau}+\frac{24}{(s\tau)^2}+\frac{54}{(s\tau)^3}+\frac{54}{(s\tau)^4}\right)\right].\end{aligned}$$

Finally we need to find the coefficients λ_n. For the sake of brevity we shall write below $\alpha_{n+\frac12}=\alpha_{n+\frac12}(\tau)$ and $\beta_{n+\frac12}=\beta_{n+\frac12}(\tau)$. It will be understood that the formulæ (A.1) and (A.2) have been introduced wherever possible. When N=2, then for insulating particles use of (73) yields

$$\lambda_1=-a_3/a_1,$$

where

$$a_1=2s\tau e^{\tau(s-2)}(3\epsilon\alpha_{\frac32}/\tau+1+2\alpha_{\frac52})+(d_{11}/d_{01})a_3,\ a_3=d_{01}(3\epsilon\beta_{\frac32}/\tau-\beta_{\frac12}-2\beta_{\frac52}).$$

Of course ν_0 and ν_1 are also readily obtained, but their explicit forms are not needed in determining the energy. For the metallic case (77) yields

$$\lambda_1=-a_{10}/a_{11}=d_{01}\beta_{\frac32}/(d_{11}\beta_{\frac32}+2s\tau e^{\tau(s-2)}\alpha_{\frac32}),$$

since, using (71), we may write

$$a_{rr}=\left(\frac{\pi}{2\tau}\right)^{\frac12}e^{-\tau}\left(\alpha_{r+\frac12}+\frac{e^{-\tau(s-2)}}{2s\tau}d_{rr}\beta_{r+\frac12}\right),$$

$$a_{rn}=\left(\frac{\pi}{2\tau}\right)^{\frac12}\frac{e^{-\tau(s-1)}}{2s\tau}d_{nr}\beta_{r+\frac12},\quad r\neq n.$$

On the other hand, the alternative solution (87) gives

$$\lambda=-\frac{a_{10}}{a^*_{11}}=-\frac{h_{10}-(a_{01}/a_{00})h_{00}}{h_{11}-(a_{01}/a_{00})h_{10}},\quad \frac{a_{01}}{a_{00}}=d_{10}\beta_{\frac32}/(d_{00}\beta_{\frac32}-2s\tau e^{\tau(s-2)}\alpha_{\frac32}).$$

Here

$$h_{00}=p_{00}/2-t_0^2,\quad h_{01}=p_{01}/2-t_0t_1,\quad h_{11}=p_{11}/2-t_1,$$

where, according to (79),

$$t_0=\left(\frac{\tau}{2}\right)^{\frac12}\left(1+\frac{e^{-\tau(s-2)}}{2s\tau}d_{00}\beta_{\frac12}\right),\quad t_n=\left(\frac{1}{2\tau}\right)^{\frac12}\frac{e^{-\tau(s-2)}}{2s\tau}d_{n0}\beta_{\frac12},\quad n\geqslant 1,$$

and from (80) and (82) and the relevant formulæ in this section

$$p_{00}=\tau\left[1+\frac{e^{-\tau(s\;\;2)}}{s\tau}d_{00}\beta_{\frac12}+\frac{e^{2\tau}}{2s}\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\}\right],$$

$$p_{01}=\frac{e^{-\tau(s-2)}}{2s}d_{10}(\beta_{\frac12}+\alpha_{\frac32}\beta_{\frac32})-\frac{e^{2\tau}}{2s^2}\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\}$$

$$+\frac{e^{-2\tau(s-2)}}{4s}\left(\frac{1}{s-1}-\frac{e^{-4\tau}}{s+1}\right),$$

$$p_{11}=\frac{\tau\alpha_{\frac32}}{3}\left(\alpha_{\frac32}+\frac{e^{-\tau(s-2)}}{s\tau}d_{11}\beta_{\frac32}\right)+\frac{e^{2\tau}}{8s^3\tau}(1-2\tau^2\overline{s^2-1})\{\mathrm{E}(2\tau\overline{s-1})-\mathrm{E}(2\tau\overline{s+1})\}$$

$$+\frac{e^{-2\tau(s-2)}}{32s^3\tau}\left[\frac{(s+1)(2s^2\tau-2\tau+5s+3)}{(s+1)^2}+5(2s\tau+1)+6\tau\right.$$

$$\left.-e^{-4\tau}\left\{\frac{(s-1)(2s^2\tau-2\tau+5s+3)}{(s+1)^2}+5(2s\tau+1)-6\tau\right\}\right].$$

When $N=2$, then for the insulating case there are five coefficients given by (73), λ_1, λ_2, ν_0, ν_1, and ν_2, of which we shall only require the first two, which have the form

$$\lambda_1=(a_2b_3-a_3b_2)/(a_1b_2-a_2b_1), \quad \lambda_2=(a_3b_1-a_1b_3)/(a_1b_2-a_2b_1),$$

where

$$a_2=(d_{21}/d_{01})a_3, \quad b_1=d_{12}\left(\frac{10\epsilon}{\tau}\beta_{\frac{5}{2}}-2\beta_{\frac{3}{2}}-3\beta_{\frac{7}{2}}\right)=(d_{12}/d_{02})b_3,$$

$$b_2=(d_{22}/d_{12})b_1+2s\tau e^{\tau(s-2)}\left(\frac{10\epsilon}{\tau}\alpha_{\frac{5}{2}}+2\alpha_{\frac{3}{2}}+3\alpha_{\frac{7}{2}}\right).$$

For the metallic case (77) yields

$$\lambda_1=(a_{12}a_{20}-a_{22}a_{10})/A=\{(d_{21}d_{02}-d_{01}d_{22})\beta_{\frac{3}{2}}\beta_{\frac{5}{2}}-2s\tau e^{\tau(s-2)}d_{01}\beta_{\frac{3}{2}}\alpha_{\frac{5}{2}}\}/A,$$

$$\lambda_2=(a_{10}a_{21}-a_{20}a_{11})/A=\{(d_{01}d_{12}-d_{02}d_{11})\beta_{\frac{3}{2}}\beta_{\frac{5}{2}}-2s\tau e^{\tau(s-2)}d_{02}\beta_{\frac{3}{2}}\alpha_{\frac{5}{2}}\}/A,$$

$$A=a_{11}a_{22}-a_{21}a_{22}=(d_{11}d_{22}-d_{12}d_{21})\beta_{\frac{3}{2}}\beta_{\frac{5}{2}}+2s\tau e^{\tau(s-2)}$$
$$\times\{d_{11}\beta_{\frac{3}{2}}\alpha_{\frac{5}{2}}+d_{22}\alpha_{\frac{3}{2}}\beta_{\frac{5}{2}}+2s\tau e^{\tau(s-2)}\alpha_{\frac{3}{2}}\alpha_{\frac{5}{2}}\}.$$

For the least square deviation method, according to (87), we obtain λ_1^* and α_2^* by replacing a_{rn}^* by a_{rn}^* (r, $n=0, 1, 2$) in the above formulæ for λ_1 and α_2, where $a_{rn}^*=a_{rn00}h-a_{0r}h_{0n}$, with

$$h_{02}=p_{02}/2-t_0t_2, \quad h_{12}=p_{12}/2-t_1t_2, \quad h_{22}=p_{22}/2-t_2^2,$$

$$p_{02}=\frac{e^{-\tau(s-2)}}{2s}d_{20}(\beta_{\frac{1}{2}}+\alpha_{\frac{5}{2}}\beta_{\frac{5}{2}})+\frac{e^{2\tau}}{16s^3\tau}(9+2\tau^2\overline{s^2-3})\{E(2\tau\overline{s-1})-E(2\tau\overline{s+1})\}$$
$$+\frac{3e^{-2\tau(s-2)}}{16s^3\tau}\left\{\frac{7}{2}+\tau\frac{s^2+1}{s^2-1}+\frac{5s-4}{(s-1)^2}-\left(\frac{7}{2}+\tau\frac{s^2+1}{s^2-1}-\frac{5s+4}{(s+1)^2}\right)e^{-4\tau}\right\},$$

$$p_{12}=\frac{e^{-\tau(s-2)}}{2s}(\tfrac{1}{3}d_{21}\alpha_{\frac{3}{2}}\beta_{\frac{3}{2}}+\tfrac{1}{5}d_{12}\alpha_{\frac{5}{2}}\beta_{\frac{5}{2}})+\frac{e^{2\tau}}{32s^4\tau^2}(9-2\tau^2 5\overline{s^2-9})$$
$$\times\{E(2\tau\overline{s-1})-E(2\tau\overline{s+1})\}$$
$$+\frac{e^{-2\tau(s-2)}}{64s^4}\left[\frac{1}{(s-1)}\left\{2\left(8s^3-\frac{9}{\tau}\right)+\frac{6s^2(11s-7)}{\tau(s-1)}\right.\right.$$
$$\left.+\frac{3(17s^3-25s^2+15s+3)}{\tau^2(s-1)^2}\right\}$$
$$\left.-\frac{e^{-4\tau}}{(s+1)}\left\{2\left(8s^3-\frac{9}{}\right)+\frac{6s^2(11s+7)}{\tau(s+1)}+\frac{3(17s^3+25s^2+15s+3)}{\tau^2(s+1)^2}\right\}\right],$$

$$p_{22}=\frac{\tau\alpha_{\frac{5}{2}}}{5}\left(\alpha_{\frac{5}{2}}+\frac{e^{-2\tau(s-2)}}{s\tau}d_{22}\beta_{\frac{5}{2}}\right)+\frac{e^{2\tau}}{64s^5}$$
$$\times\left\{\tau(11s^4-30s^2+27)-\frac{18}{\tau}(s^2-3)+\frac{81}{2\tau^3}\right\}\left\{E(2\tau\overline{s-1})-E(2\tau\overline{s+1})\right\}$$
$$+\frac{3e^{-2\tau(s-2)}}{64s^5}\left\{\frac{3(s+1)^4}{4(s-1)}\left(1+\frac{7}{2\tau(s-1)}+\frac{9}{2\tau^2(s-1)^2}+\frac{9}{4\tau^3(s-1)^3}\right)\right.$$

$$+\frac{(s+1)^2(s^2-3)}{(s-1)}\left(1+\frac{9}{2\tau(s-1)}+\frac{6}{\tau^2(s-1)^2}+\frac{3}{\tau^3(s-1)^3}\right)$$
$$+\frac{11s^4-30s^2+27}{2\tau(s-1)^2}\left(1+\frac{3}{\tau(s-1)}+\frac{3}{2\tau^2(s-1)^2}\right)$$
$$+(s^2-3)(s^2-1)\left(1+\frac{13}{2\tau(s-1)}+\frac{18}{\tau^2(s-1)^2}+\frac{9}{\tau^3(s-1)^3}\right)$$
$$+3(s+1)^3\left(1+\frac{15}{2\tau(s-1)}+\frac{45}{2\tau^2(s-1)^2}+\frac{117}{4\tau^3(s-1)^3}\right)$$
$$-e^{-4\tau}(\text{same form with } (s+1) \text{ and } (s-1) \text{ interchanged})\Bigg\}.$$

This completes the computation of all the required formulæ which occur in the first three terms of the series expansion (58) for the mutual energy of the two particles. The derivation of the higher terms can be carried out in a similar fashion, the calculations being quite straightforward although admittedly very tedious. Nevertheless, in principle, we have obtained a complete solution of the problem formulated in the introduction and have succeeded in expressing it explicitly in terms of integrals which can be readily evaluated.

In conclusion, we wish to express our appreciation to Profs. R. H. Fowler and J. E. Lennard-Jones for their interest and valuable help and suggestions. One of us (S. L.) is indebted to the Royal Society of Canada for a Fellowship enabling the work to be carried out. The other (G. P. D.) owes his thanks to the Managing Committee, Rajput College, Agra, India, for granting him leave of absence.

Summary.

Making use of the approximate Debye-Hückel equation in the theory of electrolytes a general method of determining the mutual electrical energy of two identical spherical colloidal particles has been developed. The final result for the energy is expressed as a series of which the first three terms, forming a first, second, and third approximation, are considered in detail. Both metallic and insulating particles are treated.

XV. *Thermionic Emission, Migration, and Evaporation of Barium on Tungsten.*

By J. A. Becker, Ph.D., and G. E. Moore, Bell Telephone Laboratories, New York, N.Y.*

[Received July 17, 1939.]

[Plates III.–V.]

Introduction.

In the December 1938 issue of the Phil. Mag., Dr. M. Benjamin and Dr. R. O. Jenkins described an interesting series of experiments on surface migration. While they verified Becker and Brattain's (1) results for the migration of thorium on tungsten, they conclude that barium does not migrate on tungsten as reported by Becker (2), (3). After some correspondence, Dr. Benjamin was kind enough to send us a tube of the type with which he and Dr. Jenkins studied migration of barium on tungsten. Using this tube, we have performed experiments which confirm and extend the earlier results of Becker on migration and give new information on the activation of different crystals of tungsten as a function of the amount of barium deposited.

The tube, which consists essentially of a tungsten wire 10·4 cm. long in the axis of a glass cylinder, is described in detail on p. 1056 of Benjamin and Jenkins's article. A good view of it is shown also in our photograph 3–25 (Pl. IV.). The bright line in the axis is the hot tungsten filament. The dark line just to the right is one of three vertical supports for the spirally wound grid which serves as anode. The distance between the turns of the spirals is approximately ·8 centimetre. The inside wall of the glass cylinder is coated with a fluorescent material. Barium can be vaporized from a coil located at the mouth of the side tube. The support for the tube is visible at the bottom of the picture.

General Procedure.

When we first operated the tube the emission from the filament would frequently change sporadically. These fluctuations were accompanied by sudden bursts of brilliant, but short-lived, fluorescence. This effect was especially pronounced when the anode voltage was unusually high

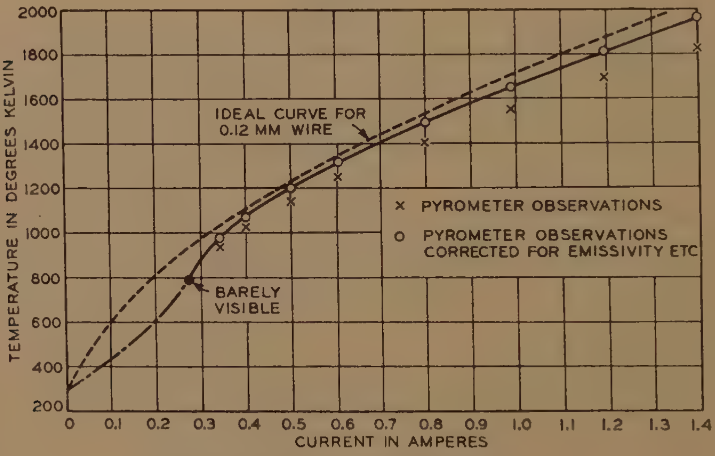
---

* Communicated by Sir O. W. Richardson, F.R.S.

and when the sides of the tube were lightly tapped. In previous work with thorium migration tubes we had ascertained that this effect was due to loose particles of fluorescent material which became positively charged and were pulled over to the tungsten filament. In the present tube these loose particles were almost completely removed by operating the filament at 1650° to 2050° K. with 4000 volts on the anode. After this treatment this spurious effect was no longer noticed if the anode potential was kept below about 2500 volts.

Barium was vaporized on to one side of the tungsten filament by heating the tantalum coil, which was coated with barium and beryllium oxides. Throughout the work here reported we used a heating current of 1·15 amperes. The temperature of the coil must have been considerably less than 1473° K., the flashing temperature used by Benjamin and Jenkins, since in our experiments the optimum activity was reached in 7 to 20 min., while in their work it was reached in about 1 min. We used a lower flashing temperature as a result of previous experience with vaporized material from oxide coated filaments (4). This experience had shown that small amounts of oxygen (or barium oxide) usually vaporize with the barium. The ratio of oxygen to barium vaporized decreases as the flashing temperature is lowered. As will be shown later, there is some evidence that a small amount of oxygen evaporated with the barium in the present tube, even at the lower flashing temperature.

Fig. 1.



Filament current, temperature curve.

In order to interpret our results clearly, it was desirable to obtain a relation between the heating current and the temperature of the filament. This relation is shown in fig. 1. The dashed curve gives the relationship for an infinitely long wire as computed from the data of Langmuir and Jones(5), for a ·12 mm. filament. The crosses give pyrometer measurements* of the temperature, and the open circles give true temperatures calculated from these observations by correcting for spectral emissivity and reflexion and absorption by the glass. The difference between these and the computed curve at *high* temperatures cannot be due to end cooling, and suggests that the diameter is slightly greater than ·12 mm. The solid point is obtained by observing the current at which the wire is just visible in a dark room. The lowest visible incandescent temperature is known to be between 775° and 800° K., so that the current at which the wire was just barely visible was plotted as 787° K. The portion of the curve between this point and room-temperature was estimated from past experience, but none of the data or conclusions in this report depends on knowledge of temperatures in this region. All references to temperature apply to the solid line curve.

Cooling due to conduction of heat along the wires is quite important in the temperature range over which migration-evaporation phenomena were studied. In a filament of this length the temperature of the centre might easily be 50° C. lower than for ideal filaments. Of course, the ends of the wire would be considerably cooler. For example, when the centre of the filament is at 1100° K. the two ends operate at 800° K., or less, for a distance of approximately 2 cm.

When the filament is heated and a potential of 500 to 2500 volts is applied to the anode grid wires the electrons from the filament strike the screen and cause it to fluoresce. From the geometry of the tube it is apparent that a small region on the filament which emits electrons will be magnified in the horizontal direction by the ratio of the tube diameter to the filament diameter. In this case the ratio is about 350. In the vertical direction there will be no magnification, but, due to the fact that the electrons are not all emitted normally to the surface, there will be a spreading out and a blurring of the image in this direction (6).

A large number of pictures of the fluorescence were taken. These pictures will be designated by two numbers: the first represents the number of the film, while the second represents the exposure on the film. The treatment preceding the picture is shown above, while the conditions under which the picture was taken are given below. The latter include the temperature at the centre of the filament and the electron emission current if different from ·1 milliampere. In most cases the pictures were taken with the side tube on the left, as shown in 3–25 (Pl. IV.). In most of these the position of the side tube appears as an indentation. In pictures 4–20 to 4–32 the camera was pointed almost directly toward the side tube.

---

* The pyrometer was sighted on the filament through a section of glass from which the fluorescent material had been removed.

Thermionic Emission as a Function of the Amount of Adsorbed Barium.

In order to interpret clearly the experiments on the migration of barium, it is desirable to know how the emission from the tungsten varies with the amount of barium on the surface. Consequently, we took two series of pictures in which the amount of barium vaporized from the barium coil on to the tungsten filament was increased in successive stages. In such a series one can take all the pictures at a constant filament temperature, or at a constant emission current. In the first case the emission current will increase, pass through a very flat optimum, and then decrease. This was the method used by Benjamin and Jenkins. In the second method the temperature required for a constant emission will decrease, pass through a very flat minimum, and will then increase. We chose this second method because of past experience.

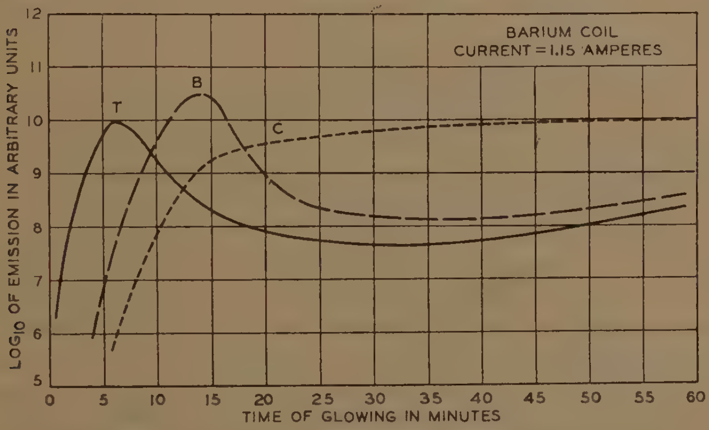
It is held generally that the emission of filaments activated by adsorbed electropositive materials is at its maximum or optimum value when the covering is approximately monatomic. Because of nonuniform rates of deposition, both along the length and around the circumference, the optimum emission vs. amount of deposit should be very flat. Calculation shows that only 1/8 as much material is deposited along an element of wire 4 cm. above or below the side arm as on a similar element directly opposite this arm. It is obvious that the rate of deposition would be a maximum around the circumference for the surfaces normal to the path of evaporated barium atoms and would decrease to zero for the points on the circumference which are tangent to the rays of evaporating atoms.

Even though the rate of deposition is uniform over some area of the filament, the emission for a given amount of deposited barium at a given temperature might still vary because of variation in the crystal structure or the work function of various regions in the area. In the past this variation was considered to be relatively small. However, the experiments with the present tube revealed that some regions require about six times as much barium as other regions to get to the optimum activity.

The two series of activation runs are shown in pictures 3–32 to 3–37 (Pl. IV.) and 4–20 to 4–28 (Pl. V.). In these pictures the emitting areas appear as bright horizontal lines. From past work in this field it is quite probable that each bright line represents the emission from a single small crystal of tungsten. On this basis we estimate that the mean crystal size was about equal to the radius of the filament, or ·006 centimetres. While it is sometimes possible to follow the activation of a single crystal, it is easier to follow the activation of larger regions composed of several neighbouring crystals.

In these pictures there are three rather distinct regions. Slightly below the side arm is a region which we may call C (centre), which is fairly well defined by a dark band in photographs 3–32, 4–22, and 4–23. Above this is a region T (top) which is bright in these three photographs; below C is a region B (bottom) which is just beginning to be bright in photographs 3–32 and 4–22. By observing both sets of photographs it can be seen that T reaches its maximum brightness after about 7 minutes of deposition and then becomes darker, while B reaches its maximum in about 14 minutes and then declines in brightness, while C continually becomes brighter and does not reach a maximum even after 45 minutes.

Fig. 2.



Postulated curves for various regions of tungsten filament.

It is certain from geometrical considerations that practically the same amount of barium falls on unit area of apparent surface of each of these regions, so that different quantities of barium are evidently required to activate individual areas to their optimum values. If it were possible to measure the activity of each of these three areas separately, the curves of the log of the emission current for the three regions against time of deposition would probably resemble those shown in fig. 2.

After picture 4–28 the filament was heated at 1070° K. for 5 minutes and photograph 4–30 taken. The activity in regions T and B increased relative to that of C and the overall activity also increased, as shown by the fact that an emission current of ·01 milliampere was obtained at 845° K. The filament was then heated to 1200° K. and 1255° K., both for 1 minute, and photograph 4–32 taken. This shows an additional increase in the activity of regions T and B relative to C, and also in the overall activity, since an emission current of ·01 milliampere was obtained at 785° K. The interpretation of these results is clear from fig. 2, since the evaporation or migration of barium from regions T and B resulting from such treatment should cause the emission of these regions to return toward their optima.

From the times of barium deposition noted above the pictures, it can be seen that region T requires only about 1/6 the amount of barium to develop optimum emission as is required for region C. Two interpretations seem possible to account for this behaviour. The first is that region C is very rough, region B is less rough, and region T is quite smooth by comparison. A longer evaporation time would then be required to cover unit area of true surface in region C with a given number of atoms than for the other regions. The second interpretation is that some regions require about six times as much barium per cm.^{2} of true surface as other regions to reach their optimum emission.

Barium Migration.

Three preliminary runs were made during which barium from the coil was vaporized on to the tungsten filament and the subsequent behaviour studied. No photographs were made in these runs, but each gave satisfactory evidence of migration and the range of conditions over which migration could be expected. Subsequent runs were photographed, and some of these photographs are reproduced. Unless noted, all photographs were made with emission currents of ·1 milliampere and migration was allowed to proceed without anode voltage.

The photographs begin with 1–15. The filament had been cleaned of barium by glowing at 1890° K. for 3 minutes. With the filament cold, the barium generating coil (Ba coil) was operated for 39 minutes at 1·15 amps. Photograph 1–15 was then taken. Between 1–15 and 1–20 the filament was glowed at 970° K. for 102 minutes. The activity of the filament had increased as a result of the glowing and some migration is shown, since a few streaks of fluorescence are now visible to the right of the dark upright wire. The temperature was then raised successively to 1025° K., 1070° K., and 1135° K. for the times indicated on photographs 1–24, 2–1, and 2–5 respectively. The activity did not suffer from this glowing until the temperature exceeded 1070° K., when it fell somewhat, as shown in 2–5. But the photographs show that migration had occurred during operation at 1025° (1–24) and even more during operation at 1070° (2–1), when the fluorescence extended clear around the tube. This result indicates that the barium had migrated to the back of the filament. The fact that the standard emission, ·1 milliampere, could be obtained at the same temperature (970° K.) as in 1–15, indicates that evaporation is not very serious at these temperatures, although evaporation had apparently occurred on region C. After photograph 2–5 was taken the barium coil was again operated for 33 minutes. The tungsten filament was then heated at 1070° K. for 1 hour and photograph 2–20 was taken. This shows fluorescence completely around the filament on several vertical zones, which is conclusive evidence that migration had occurred at 1070° K. The fact that the filament gives the standard emission current at 925° K. indicates that evaporation was not very important at 1070° K. Other photographs were taken of similar cycles during which the tungsten was cleaned off, coated with barium on one side, and the barium made to migrate to the other side. These photographs are not reproduced, because of the space required.

However, one additional mode of migration is worthy of some emphasis. After making photograph 2–20 the filament was operated at progressively higher temperatures during which evaporation was predominant. Photographs 3–1, 3–2, 3–7, 3–10, 3–13, and 3–20 show the results of this operation. The filament must be operated at increasingly higher temperatures to obtain the standard emission, which is now chiefly from the ends and, in 3–10 to 3–20 inclusive, predominantly from the upper end *. Since this was the negative end of the filament it suggested that the migrating barium was positively charged. A similar result had been reported by Johnson (6) earlier for the case of thorium on tungsten. To test this the filament was cleaned off, its polarity reversed, and a similar deposition and heating cycle was again followed. The end results for this second cycle are shown in photographs 4–5 and 4–9. These photographs show that the emission on the bottom is much stronger than before. In fact, 4–9 indicates that most of the emission comes from the bottom sections. The photographs are thought to be evidence that at least some of the barium present on the filament is positively ionized, so that it is impelled toward the negative terminal of the filament. It should be noted that the fluorescent rings due to barium in these latter photographs go completely around the tube, thus showing migration.

If one accepts the above, or other, evidence for ionization of some of the barium present, it would be very surprising not to observe barium migration, since the fields between ions would have a strong tendency to aid the concentration gradients and force the barium to migrate.

Regions T, C, and B lose their barium rapidly compared to a region R (rear) just on the other side of the filament. This is shown in photographs 2–33, 3–1, and 4–2 and 4–3. The barium on R is received by migration from T and C. Since region R is covered close to the optimum

---

* In 3–20 to 3–30 the general illumination is due to light from the filament. a good deal of which is scattered by the fluorescent coating.

value as judged by the brilliance of its fluorescence, migration from regions T and C would have to occur against a concentration gradient if we adopt the " roughness " interpretation offered previously. It is easier to believe that regions C and T have several layers of barium present and that this then migrates to region R as a result of thermal agitation and of surface fields produced by concentration gradients of neutral and ionized barium. An examination of photographs 4–2 and 4–3 will show that region R has retained its activity even after glowing for 5 minutes at 1260° K., and that regions T, C, and B have already lost their activity at 1135° K. This shows that evaporation rates are quite different for different regions of the filament. From the pattern and activity shown in 4–9 it follows that considerable barium is still present on the centre of the filament after glowing at 1540° K. for 47 minutes.

The observations of migration and evaporation on different regions of the tungsten are summarized in the following table :—

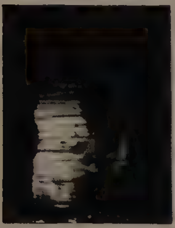
	Region.			
	T.	B.	C.	R.
	°K.	°K.	°K.	°K.
Migration begins	1025	1025	970	—
Migration is rapid	1070	1070	1025–1070	—
Evaporation detected	1070	1070	1025	1260
Evaporation is pronounced	1140	1100–1140	1070	1310

Migration of barium on tungsten has been observed elsewhere. Muller(7) has made some beautiful experiments in which migration was observed over smaller distances than those reported here. We have verified these results in a duplicate of Muller's tube. Dr. R. P. Johnson, of the American General Electric Company, informed us that he observed migration of barium on tungsten.

Evidence for Oxygen Contamination.

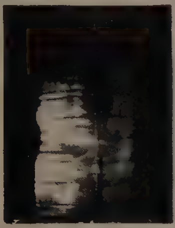
Several measurements were made of the clean tungsten emission of the filament. In order to clean the tungsten it was necessary to operate the filament for several minutes at temperatures of the order of 2000° K. or above. The observations of clean tungsten emission are shown in fig. 3 by small open circles. The straight line shows a calculated curve for clean tungsten, using the emission data of Langmuir and Jones (5). The temperatures are shown in the upper part of the figure. At these temperatures the effect of end cooling would not be very pronounced, and the agreement of the calculated and observed values is seen to be

1-15 1.15 AMP.,
BA 39 MIN.



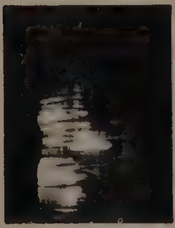
970°K

1-20
970°K 102 MIN.



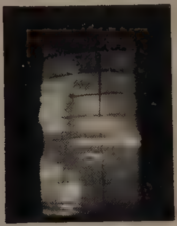
950°K

1-24
1025°K 55 MIN.



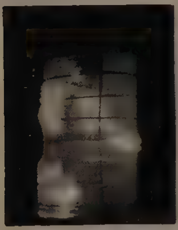
970°K

2-1
1070°K 17 HRS.



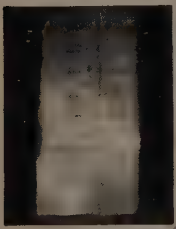
970°K

2-5
1135°K 57 MIN.



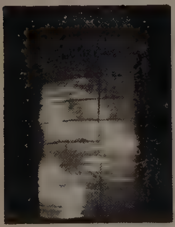
990°K

2-20
1070°K 60 MIN.



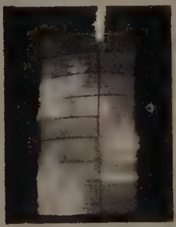
925°K

2-27
1095°K 74 MIN.



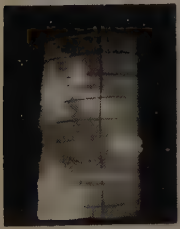
970°K

2-33
1095°K 16 HRS.



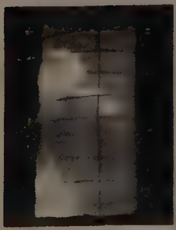
1025°K

3-1
1260°K 3 MIN.



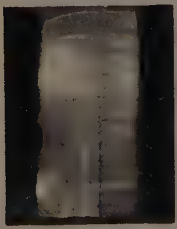
1040°K

3-2
1310°K 3 MIN.



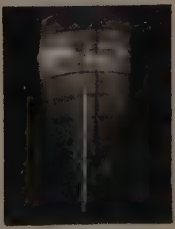
1135°K

3-7
1360°K 3 MIN.



1210°K

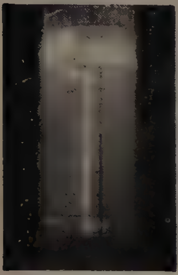
3-10
1410°K 5 MIN.



1310°K

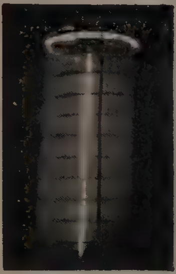
BECKER & MOORE. Phil. Mag. Ser. 7. Vol. 29. Pl. IV.

3-13
TUBE MOVED



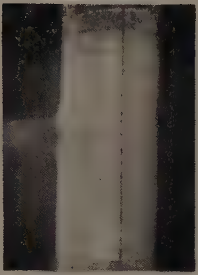
1310°K

3-20
1540°K 3 MIN.



1540°K $I_b=.04$

3-25
1820°K 8 MIN.

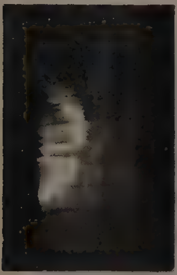


1965°K $I_b=.14$

3-32
BA 5 MIN.

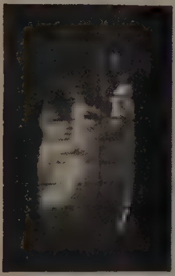
910°K $I_b=.01$

3-33 10 MIN.
BA TOTAL



900°K $I_b=.01$

3-34 15 MIN.
BA TOTAL



900°K $I_b=.01$

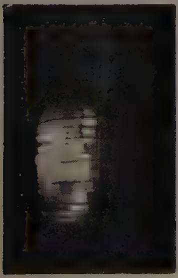
3-35 20 MIN.
BA TOTAL

885°K $I_b=.01$

3-36 40 MIN.
BA TOTAL

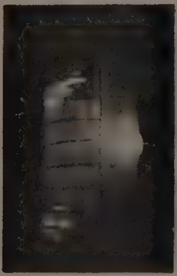
900°K $I_b=.01$

3-37 59 MIN.
BA TOTAL



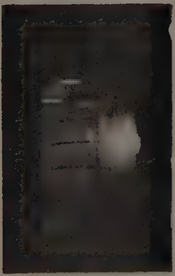
910°K $I_b=.01$

4-2
1135°K 16 HRS.



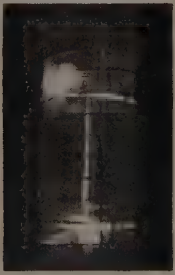
910°K $I_b=.01$

4-3
1260°K 5 MIN.



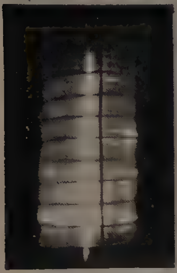
870°K $I_b=.01$

4-5
1410°K 5 MIN.



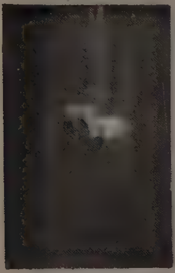
1225°K $I_b=.01$

4-9
1540°K 37MIN.



1465°K I_b=.01

4-20
BA 4MIN.



1055°K I_b=.002

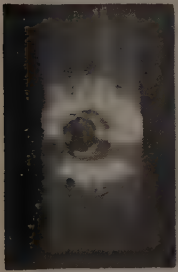
4-21 6MIN.
BA TOTAL

945°K I_b=.01

4-22 8MIN.
BA TOTAL

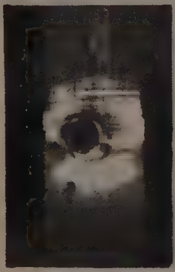
900°K I_b=.01

4-23 10MIN.
BA TOTAL



915°K I_b=.01

4-24 12MIN.
BA TOTAL

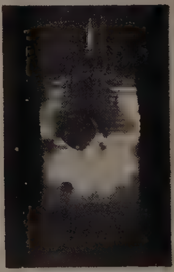


900°K I_b=.01

4-25 14MIN.
BA TOTAL

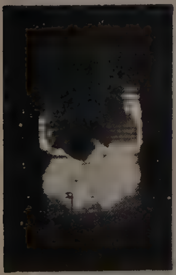
875°K I_b=.01

4-26 16MIN.
BA TOTAL



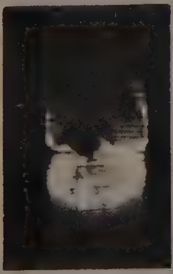
875°K I_b=.01

4-27 20MIN.
BA TOTAL



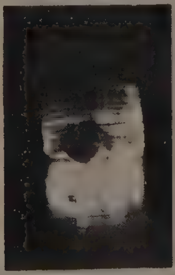
875°K I_b=.01

4-28 25MIN.
BA TOTAL



875°K I_b=.01

4-30
1070°K 5MIN.



845°K I_b=.01

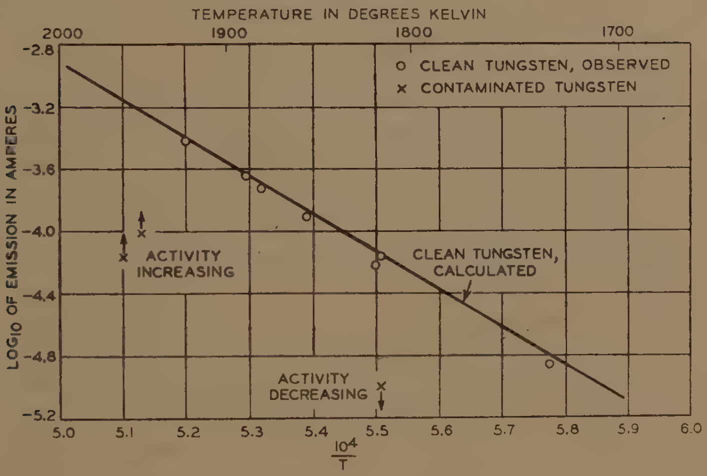
4-32
1200°K 1MIN.
1255°K 1MIN.

785°K I_b=.01

good. These observed emissions were quite steady and reproducible, even after repeated glowing at high temperatures.

After barium had been deposited on to the tungsten and the tungsten was heated, the barium evaporated and the emission decreased. But instead of decreasing to the emission value corresponding to that for clean tungsten, the emission of the filament decreased well below that for clean tungsten for temperatures between 1700° K. and 1850° K. When the filament was then operated at temperatures above 1850° K. the activity increased to that of clean tungsten, and subsequent operation at still higher temperatures produced no further changes in activity. This is shown by the three crosses in fig. 3, where emissions of only 1/7 to 1/10 that of clean tungsten were observed.

Fig. 3.



Emission from clean and contaminated tungsten in Benjamin's electron microscope tube.

It is well known that, for any covering of barium on tungsten, the emission is greater than that for clean tungsten. Since these three points and many similar observations show emissions less than those for clean tungsten, it is very probable that the metal was contaminated with some impurity other than barium, presumably oxygen. It is known that oxygen contamination causes the activity of clean tungsten to decrease and that the contaminating oxygen evaporates rapidly at temperatures of 1900° K. or above. These facts suggest that the contamination is due to oxygen.

It is probable that the oxygen evaporated from the Ba coil either as oxygen gas or as BaO. Any adsorbed barium oxide would probably be reduced by hot tungsten to adsorbed barium and adsorbed oxygen. It is quite probable that the rate of migration of barium would depend on the amount of oxygen contamination and might decrease as the amount of oxygen increases. This may be the reason why Benjamin and Jenkins did not observe migration.

Theory of Migration.

From the Maxwell-Boltzmann law, it follows that when the temperature of a surface is high enough so that a few of the adsorbed atoms or ions are able to evaporate, there must be an enormously greater number which leave the surface for a short distance and are then pulled back to it by surface forces. It is quite improbable that they will strike the surface at the same point at which they left it; hence one would expect migration to occur whenever evaporation occurs. One would also expect that some migration would occur even though the temperature is so low that the rate of evaporation is negligibly small. One would also expect that migration can be observed in a temperature range slightly below that at which evaporation is appreciable. Since the evaporation rate is a function of concentration of adatoms as well as temperature (2), (3), one would expect that migration would depend not only on concentration gradient and temperature, but also on the concentration itself. There may also be some migration in which the atoms do not leave the surface; but, of course, it is difficult to say how far an atom must go before it leaves the surface temporarily. If, in this *incipient evaporation*, some of the adsorbed particles act like ions, one would then expect that they would be affected by a field parallel to the surface. This, we believe, is the explanation for the preferred migration toward the negative end of the filament.

Summary of Experimental Results.

When barium is deposited on tungsten, the thermionic activity of the tungsten increases, comes to a maximum, and then decreases. It has frequently been found by emission measurements that this optimum corresponds to about a monomolecular layer. However, data obtained in this work show that some regions of the filament require more than five times as much barium as others for optimum emission.

Photographs are presented which show that the rates of both migration and evaporation depend on the crystal surface, the temperature, and amount of barium on the surface. Barium migration on tungsten can be observed at temperatures as low as 970° K., is readily observed at 1025° K., and is rapid at 1070° K. Evaporation is observed on some crystals at temperatures as low as 1025° K., while on others it is slow even at 1260° K. At 1300° K. it is rapid for all crystals. These temperatures probably vary with the oxygen contamination which comes over to the filament with the barium. For barium concentrations near the optimum there exists a range of temperature over which migration is readily observed, but where evaporation is not noticeable.

Measurements of electron emission after all the barium is evaporated show that the filament was contaminated by an electronegative material, probably oxygen.

Barium tends to migrate toward the negative end of the filament, thus indicating ionization of adatoms.

We take pleasure in thanking our colleague, Dr. A. J. Ahearn, for several suggestions and help in taking some of the data.

References.

(1) J. A. Becker and W. H. Brattain, Phys. Rev. xliii. p. 428 (1933).
(2) J. A. Becker, Trans. Am. Electrochem. Soc. lv. p. 153 (1929).
(3) J. A. Becker, Trans. Faraday Soc. xxviii. p. 148 (1932).
(4) J. A. Becker, Phys. Rev. xxxiv. p. 1323 (1929).
(5) Langmuir and Jones, G. E. Review, xxx. p. 310 (1927).
(6) R. P. Johnson, ' Journal of Applied Physics,' ix. p. 508 (1938).
(7) E. W. Muller, *Zeits. f. Phys.* cvi. p. 541 (1937) ; cviii. p. 668 (1938).

XVI. *The Radiation of the Acoustic Air-Jet Generator derived from direct Observation of the Amplitude of the Aerial Vibrations in the Oscillator.*

By JUL. HARTMANN and FREIMUT LAZARUS *.

[Plates VI.–VIII.]

[Received February 2, 1939.]

Introduction.

THE acoustic air-jet generator was described in the present periodical some years ago †. So it will suffice briefly to remind the reader of the principle. An air-jet with a velocity exceeding that of sound is discharged from a container through a nozzle N (fig. 1). In such a jet there is a series of equidistant intervals a_1b_1, a_2b_2 . . . , the intervals of instability. If a cylindrical oscillator O, *i. e.*, a cylindrical bore with a sharpened edge made in the end of a piece of brass, is arranged with its aperture turned against the stream, the air in the oscillator will vibrate if, and only if, the aperture of the oscillator is within one of the intervals of instability. So, with the oscillator thus arranged, the nozzle, jet, and oscillator constitute an acoustic generator radiating waves in the surrounding medium. The generator is particularly well suited for the production of waves within the frequency range 10,000 to 100,000 cycles/sec. The intensity of the waves is comparatively very great. If the excess-pressure in the container is taken to be 3·2 kg./cm.^{2}, the acoustic output will for instance be 6 watts at 50,000 cycles/sec. and about 160 watts at 10,000 cycles/sec. The efficiency is in both cases about 4 per cent., *i. e.*, the acoustic output is 4 per cent. of the power required, theoretically, for maintaining the jet. It is here assumed that the dimensions d_0, d, and l in fig. 1 are all of the same size, the accepted rule in the design of this generator, seeing that with $d_0 = d = l$ a maximum of efficiency is attained.

The performance of the acoustic air-jet generator has recently been made the subject of comprehensive investigations. One of the results of this research work is a formula for the total radiation of acoustic energy derived from measurement of the distribution of intensity round the generator by calibrated Rayleigh disk arrangements. The formula referred to is

---

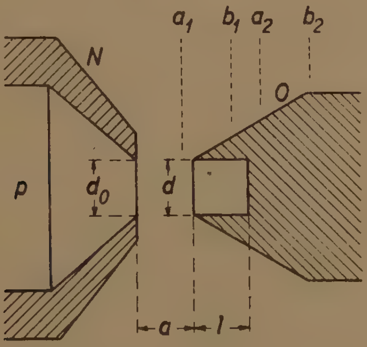
* Communicated by the Authors.
† Phil. Mag. xi. p. 926 (1931).

$$I_0 = 3{\cdot}0\,d^2\sqrt{p-0{\cdot}9}\ \text{watt}\ (d\ \text{mm.},\ p\ \text{atm.}),$$

where $d=d_0=l$ is the diameter of the oscillator and p the excess-pressure in the container. The formula holds good with the oscillator within a certain part of the first interval of instability—just behind the frontier facing the nozzle—where the efficiency is a maximum. It was derived from the determination of the total radiation in 14 cases with values of d between 2 and 6 mm. and of p between 2·07 and 3·44 atm. The standard deviation of the numerical coefficient from the average 3·0 was 13 per cent.*

Now, as will be gathered, the "photometric" measurement of the total radiation is a rather intricate affair. This consideration gave rise to the investigations reviewed in the following sections, and having for their aim the measurement of the total radiation on the basis of direct observation of the amplitude of the aerial vibrations in the oscillator.

Fig. 1.



Main members of the acoustic air-jet generator.

The Aerial Vibrations in the Oscillator made visible by means of the Riemann Mirror. Riemann Mirror Oscillograms.

If a jet with a velocity greater than that of sound impinges perpendicularly on a wall, an accumulation of air is produced in front of the wall. The frontier of this accumulation facing the stream is extremely sharply defined. One may look at the accumulation as a stationary

* The complete account of the investigation on the performance of the air-jet generator, together with others on the constructional development, the techniques of measurement, the *modus operandi*, etc., of this generator, will appear in English in the course of a few months.

compression wave in the jet. This wave has been studied by Riemann, and so we may use the name a Riemann-wave for it. In the front of this wave all the characteristics of the jet, pressure, density, temperature, velocity, change quite suddenly, *i. e.*, within a layer the thickness of which is estimated to be of the order of magnitude of the free path of the molecules. If, for instance, the pressure in front of the layer is 0·6 kg./cm.^{2}, it may be 2·4 kg./cm.^{2} behind the layer. From this it will be gathered that the front of the Riemann-wave must act as a most brilliant mirror. In the following this mirror is spoken of as the Riemann mirror. Fig. 2 *a*, *b* (Pl. VI.) shows two edgewise illuminated Riemann mirrors in front of a plane wall. The jet proper is discharged from the nozzle seen to the left; it does not appear in the photos The jet was the same in the two cases, but the distance from the nozzle to the wall is different. This accounts for the different shapes of the mirror. Fig. 3 *a* (Pl. VI.) shows a Riemann mirror as a bright disk in the space before a plane surface (to the right), while in fig. 3 *b* (Pl. VI.) the reflexion in the mirror of two brightly illuminated pin-holes in a piece of cardboard is seen.

Now, it is a fact of great interest that a Riemann-wave and thus a Riemann mirror is also formed in front of the oscillator of the acoustic air-jet generator. Fig. 4 *a* (Pl. VI.) shows the illuminated mirror when the oscillator (to the right) is still outside the first interval of instability, *i. e.*, between the same and the nozzle (to the left). Fig. 4 *b* (Pl. VI.) shows what happens when the oscillator is carried inside the interval. The sharp bright line of the stationary mirror is replaced by a band of some width. If the oscillator is carried farther into the interval the width of the band increases very considerably and rather soon, fig. 4 *c* (Pl. VI.), reaches a size very nearly equal to that of the aperture of the oscillator, which in the case considered was 6 mm.

Quite obviously it is the vibrations of the air in the oscillator which are brought to view in the photographs, figs. 4 *b*, *c* (Pl. VI.). In order to make quite sure of this, the illuminated Riemann mirror was projected on to a wall with a narrow slit perpendicular to the mirror. Behind this slit a photographic film was moved at a speed of 10 m./sec. in a direction perpendicular to the slit. In this way oscillograms of the vibrating front of the air in the oscillator were produced. Figs. 5 *a–c* (Pl. VII.) show three such oscillograms. They originate from a generator of $d_0=d=l=6$ mm., operated at an excess-pressure of 2·84 atm. Fig. 5 *a* (Pl. VII.) shows ultra-rapid vibrations produced with the oscillator just at the entrance to the first interval of instability. In fig. 5 *b* (Pl. VII.) the normal vibrations of a frequency of about 10,000 cycles per sec. are seen. Finally fig. 5 *c* (Pl. VII.) shows vibrations with the oscillator in the second interval of instability.

Measurement of the Total Radiation by means of the Riemann Mirror.

We pass on to explain the idea of calculating the total radiation of the generator from the observed amplitude A of the vibrations of the Riemann mirror, which vibrations are assumed to be identical with the maximum vibrations of the air in the oscillator. We shall visualize these latter vibrations as produced by a continuous train of waves being created at the orifice of the oscillator, running down to the bottom of this oscillator, being reflected from the same, and on arrival at the orifice being emitted into the surrounding space constituting the radiation of the generator. This view is, of course, very far from picturing the actual *modus operandi*, but it may form the basis of a theory for the calculation of the radiation. Denoting by a the amplitude of the train of waves referred to, by ρ the average density of the air in the oscillator, by d the diameter of the oscillator-bore, by N the frequency of the vibrations, and by c the velocity of sound in the medium outside the generator, the total radiation is, on this assumption, determined by

$$\mathrm{I}=\frac{1}{2}\rho\cdot\frac{\pi}{4}d^2.(2\pi\mathrm{N}a)^2c.10^{-7}=0{\cdot}0528\rho d^2\mathrm{N}a \text{ watt}, \quad . \quad . \quad . \quad (1)$$

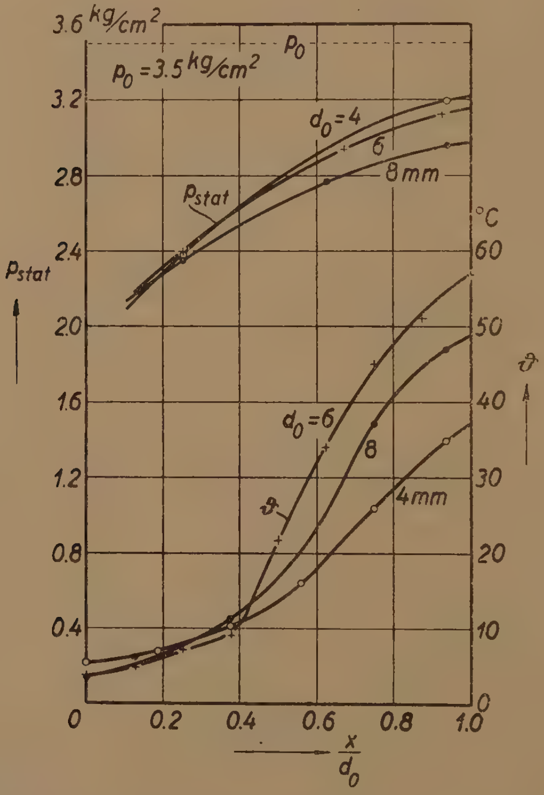
where again $a=\frac{\mathrm{A}}{2}=\frac{\mathrm{B}}{4}$, if B is the breadth of the band drawn by the vibrating Riemann mirror. We shall calculate the average density ρ of the air in the oscillator from the formula

$$\rho=\frac{0{\cdot}001293}{1+0{\cdot}00367}\cdot p, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

by introducing for p the average absolute pressure and for the average temperature in the oscillator.

In order to determine the latter two quantities a special apparatus, shown in fig. 6 (Pl. VIII.) was constructed. It consists of an acoustic generator of standard design, apart from the oscillator, which was furnished with a narrow bore through the bottom so that a slender pressure gauge or a thermo-couple could be carried into the oscillator from behind. The gauge or couple was carried by a three-coordinate slide arrangement in order to facilitate the centreing and the axial displacement. The position of the tube with the nozzle could be adjusted relatively to the oscillator by means of the hand-wheel seen to the left in the figure. With this apparatus, the details of which need not be described in this place, the curves in fig. 7 were produced. The abscissa is the distance x from the orifice of the oscillator measured with $d_0(=d=l)$, the joint diameter of the jet-hole and the oscillator-bore, as unit. The uppermost set of curves shows the distribution of the static pressure along the axis of three oscillators of $d=l=4$, 6, and 8 mm. respectively. The lower curves give the distribution of the temperature. One might expect the curves in each group to coincide. They do so substantially, though marked differences are manifest. Fig. 8 shows other experiments where the absolute pressure p_0 in the container was taken as parameter. From

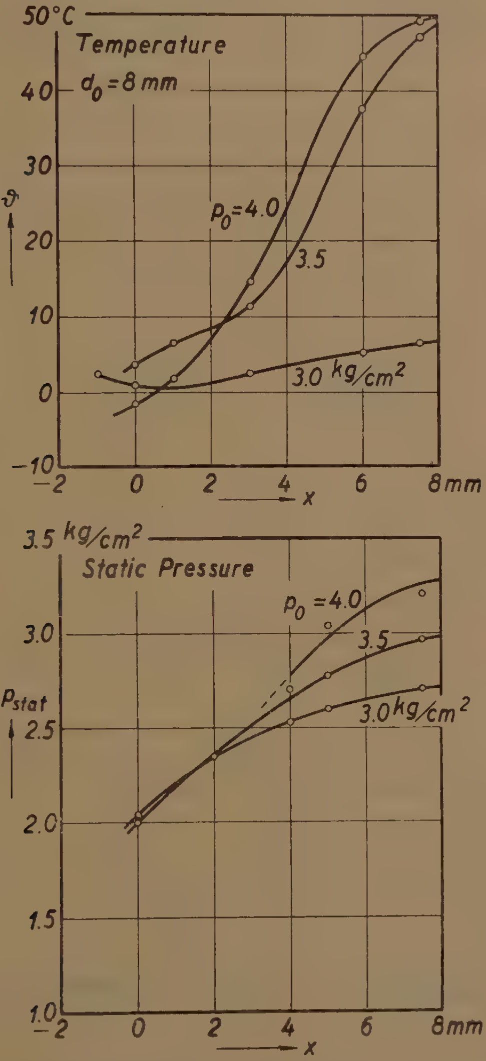
Fig. 7.



Distribution of the static pressure and the temperature in the oscillator during operation.

the whole material the values $p=0{\cdot}78\,p_0$ and $\vartheta=20°$ C. were derived for the average pressure and the average temperature over the interior of the oscillator during operation. These values were then used in (2) for the calculation of ρ. For the observation of the amplitude *a* a special optical bench, fig. 9 (Pl. VIII.), which cannot be described in this place, was designed. With this bench the amplitude, or rather the breadth B, of the band drawn by the Riemann mirror was directly observed on a scale. It turned out that B, and so a, is very nearly constant within that part of

Fig. 8.

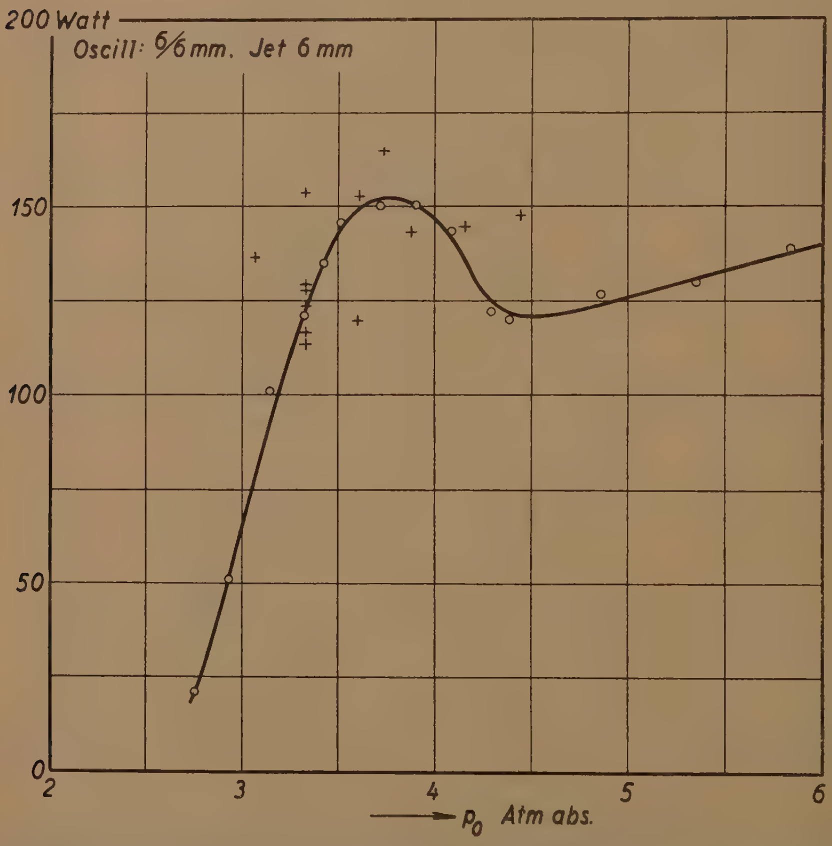


Distribution of the static pressure and the temperature in the oscillator during operation.

the first interval of instability for which the formula $I_0 = 3{\cdot}0\, d^2\sqrt{p - 0{\cdot}9}$, derived from the Rayleigh disk measurement, was found valid. In fig. 10

the points marked by small circles represent the values found by amplitude measurement according to the Riemann mirror method. The curve is drawn through these points. The points marked by crosses represent values of the total radiation as found from the Rayleigh disk investigations, *i. e.*, the values from which the formula for I_0 was derived. As, however, the amplitude observations were all taken at an oscillator of 6 mm. aperture, while the Rayleigh disk investigations were carried out with oscillators of various openings, the radiations calculated from the latter investigations had all to be reduced to a diameter value of 6 mm. As, however, it is a well established fact that I_0 is proportional to d^2, this reduction is quite legitimate. The points marked by crosses may safely

Fig. 10.



Values for the total radiation determined by the Rayleigh disk, the crosses, compared with values computed from amplitude observations according to the Riemann mirror method, the circles.

be considered as actual observations. So fig. 10 shows that the amplitude measurement may serve to determine the radiation, a fact of considerable practical interest since this measurement is of a very simple character, requiring only comparatively uncomplicated apparatus.

The authors are indebted to the Trustees of the Carlsberg Foundation, the Rask-Φrsted Fund, and the Otto Mönsted Fund for financial aid in connexion with the completion of the investigations reviewed above.

The Laboratory of Technical Physics,
The Royal Technical College,
Copenhagen.

XVII. *On the Theory of Light-scattering.*

By S. Parthasarathy, D.Sc., F.Inst.P.
(Nobel Institute, Stockholm) *.

[Received July 6, 1939.]

A paper with the above title, by Hans Mueller, appeared in the Proceedings of the Royal Society, putting forward a new theory based on the assumption that "frozen-in" strains do also scatter light. The theory is very interesting, but seems to be wholly based on the experimental work of R. S. Krishnan on light-scattering in binary liquid mixtures, colloids, and glasses. Based on Krishnan's experimental work are also some other interesting papers, one of which is that of J. Frenkel, who postulates "the emulsoidal nature of binary liquid mixtures at critical solution temperature and well above it," and who also suggests another definition for the critical solution temperature. R. Gans's paper tries to explain the anomalous results of Krishnan by assuming the existence of clusters of molecules in such systems. But his own theory leads one to the conclusion, even in most favourable cases, that one detects clusters of disk-shape of diameter 8000 Å.U. and over by employing incident light of 4000 Å.U., which is therefore not the case contemplated in Gans's paper. Further, since there has been no confirmation by any worker of Krishnan's results, and consequently very little support for the above theories, this paper critically examines the results obtained by Krishnan. Indeed, as it will appear later, this paper makes it clear that no new theory such as those mentioned above is called for.

It is well known that in the case of molecular light-scattering, usually called the Rayleigh scattering, by gases and vapours, the depolarization of transversely scattered light gives much valuable information regarding the anisotropy of the molecule. For liquids, in addition to the above facts, one gets a fair idea also of the polarization field surrounding the molecules. This is so when the incident light is unpolarized. For incident light with electric vector vertical the relationship between ρ_{u_0} and ρ_{v_0} is given by †

$$\rho_{u_0}=\frac{2\rho_{v_0}}{1+\rho_{v_0}},$$

---

* Communicated by Prof. R. H. Fowler, F.R.S.

† In this paper we adopt the older and unambiguous definition for ρ, the depolarization factor, as $\frac{\text{Horizontal component}}{\text{Vertical component}}$, and *not* as $\frac{\text{weaker component}}{\text{stronger component}}$, which is very misleading. Further, in the latter case one has to say which component is weaker, as otherwise it is meaningless.

the subscript u_0 or v_0 denoting the nature of the incident light, unpolarized or vertical. But for incident light with electric vector horizontal ρ_{h_0} is always unity both for gases and pure liquids, except for colloids and impurities, occluded or suspended, in which cases ρ_{h_0} is greater than one.

It would be worth while to examine more closely how and where ρ_h can be greater than unity, as obtained by Krishnan, for substances other than colloids and impurities.

One must remember that the above relationship holds good only if the incident light is parallel and not convergent. But the results become modified in the latter case, and Gans, in a notable paper, has derived the exact values of the depolarization if one knows the aperture of the lens used for condensing the incident beam. His results are given below:

$$\rho_v = \rho_{v_0} + \tfrac{1}{4}\theta^2,$$

$$\rho_h = 1 + \frac{1}{\rho_{v_0}} \cdot \frac{1}{4}\theta^2,$$

and

$$\rho_u = \rho_{u_0} + \tfrac{1}{2}\theta^2.$$

Here ρ_{v_0} and ρ_{u_0} are the values of depolarization for the incident parallel light, *i. e.*, the ideal case contemplated in the theory of light-scattering, and θ is the semi-angle of convergence of the lens used to focus the incident light on gas or liquid. It will be seen from these formulas that ρ_h is always greater than unity, being more so for large angles of convergence as well as for low values of ρ_v.

In this paper I have taken for examination one detailed paper of Krishnan; this treatment applies to the others as well. This paper is also representative, for the reason that the authors I have mentioned earlier had depended for their theories largely on this.

The following tables contain results taken from Krishnan's paper *. There is nothing new we have added except the figures in the last column, which are taken from the graph. These are found to give the most likely aperture of the lens.

The last four tables give the values of ρ_v and ρ_h for only four values of the aperture of the lens, and these have been obtained by the use of Gans's formula. Only four apertures have been taken, as they are found to cover all the mixtures studied by Krishnan.

The diagram shown on p. 152 gives the results concisely. The continuous curves in the diagram represent the theoretical relation between ρ_v and ρ_h for different apertures as given by Gans's formula. The dots represent the values for these mixtures obtained experimentally by Krishnan.

---

* All the depolarization factors in his paper have been converted to the definition adopted here.

TABLE I.

Experimentally Determined Values of Krishnan *.

ρ_v %.	ρ_h.	Aperture.
Mixture (*a*) : Carbon disulphide and methyl alcohol.		
0·27	4	F/4·5
0·34	3·7	
0·62	2·2	
0·93	1·70	
1·1	1·47	
1·3	1·37	
1·7	1·30	
2·1	1·11	
2·8	1·08	
4·1	1·0	
Mixture (*b*) : Methyl alcohol and (*n*) hexane.		
0·37	3·7	F/4·7
0·37	3 ?	
0·62	1·88	
0·76	1·67 ?	
0·76	1·6	
0·84	1·47	
1·0	1·24	
1·1	1·15	
1·4	1·0	
Mixture (*c*) : Water and isobutyric acid.		
0·34	3·7	F/4·7
0·5	2·56	
0·6	1·92	
0·7	1·7	
1·1	1·43	
1·3	1·23	
1·5	1·1	
1·8	1·0	
Mixture (*d*) : Aniline and cyclohexane.		
0·12	5·9 ?	F/6·25
0·15	4·5	
0·27	2·5	
0·37	1·75	
0·43	1·47	
0·63	1·4	
0·93	1·3	
1·2	1·2	
1·4	1·15	
1·5	1·11	
2·0	1·08	
2·5	1	
Mixture (*e*) : Nitrobenzene and (*n*) hexane.		
0·12	6·25 ?	F/4·5
0·27	4	
0·37	3	
0·5	2·4	
0·93	1·64	
1·3	1·5	
2·0	1·33	
2·3	1·2	
3·0	1·15	
4·1	1	
Mixture (*f*) : Aniline and (*n*) hexane.		
0·19	4·8	F/5·6
0·32	2·67	
0·32	2·2 ?	
0·43	2·0	
0·49	1·70	
0·62	1·43	
0·93	1·28	
1·0	1·19	
1·1	1·11	
1·4	1	

---

* The dots in the graph represent these values.

TABLE I. (*cont.*).

ρ_v %.	ρ_h.	Aperture.	ρ_v %.	ρ_h.	Aperture.
Mixture (g) : Phenol and water.			Mixture (h) : Water and triethylamine.		
0·07	8·3 ?	F/5·6	0·38	3	F/4·5
0·12	6·25		0·85	2·4	
0·19	3·9		2·7	1·8	
0·23	3·3		5·3	1·6	
0·27	2·6		8·5	1·5	
0·34	2·13		9·3	1·43	
0·43	1·96		14·0	1·3	
0·6	1·64		17·0	1·2	
0·7	1·5				
1·0	1·3				

TABLE II.

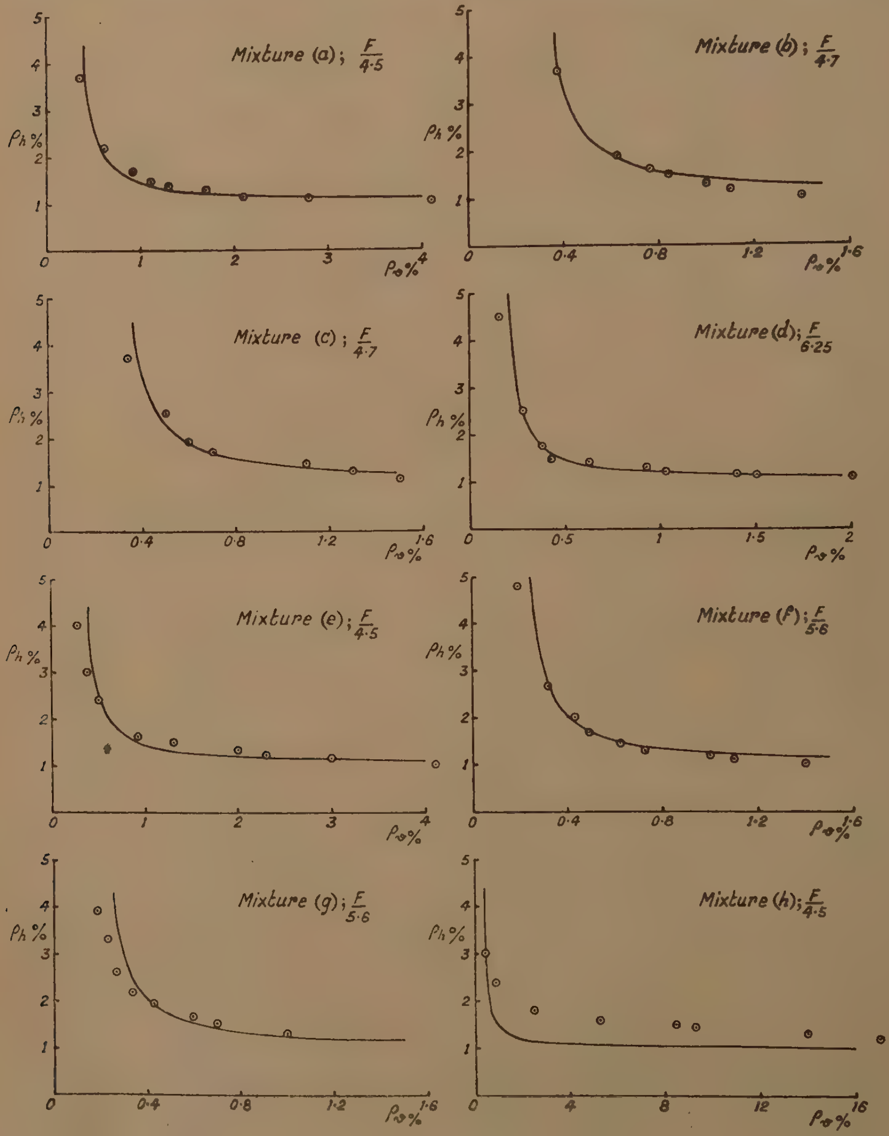
ρ_v and ρ_h from Gans's Theory for Different Apertures *.

Aperture.	ρ_v %.	ρ_h.	Aperture.	ρ_v %.	ρ_h.
F/4·5	0·40	4·4	F/4·7	0·36	4·5
	0·51	2·55		0·40	3·3
	0·61	2·03		0·44	2·75
	0·71	1·78		0·48	2·4
	0·81	1·62		0·52	2·17
	1·01	1·44		0·56	2·0
	1·31	1·31		0·68	1·7
	1·61	1·24		0·88	1·47
	1·81	1·21		1·08	1·35
	2·01	1·18		1·48	1·23
	2·31	1·16			
	2·81	1·12			
	3·41	1·10	F/6·25	0·24	3·0
	4·01	1·09		0·36	1·8
F/5·6	0·26	4·3		0·46	1·53
	0·27	4·0		0·66	1·32
	0·29	3·2		0·86	1·23
	0·30	3·0		1·06	1·18
	0·40	2·0		1·36	1·13
	0·60	1·5		1·56	1·12
	0·90	1·3		1·76	1·10
	1·0	1·25		1·96	1·09
	1·2	1·20		2·16	1·08
	1·5	1·17			
	1·8	1·13			
	2·0	1·11			

---

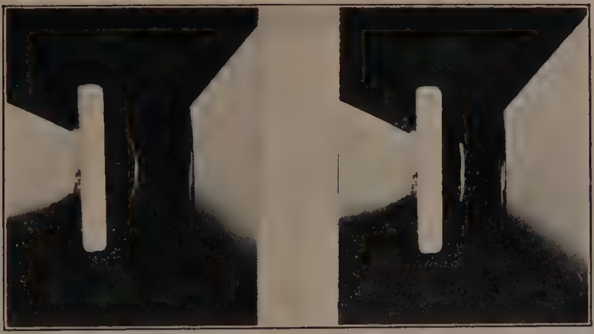
* The continuous lines represent these values in the graph.

All the dots lie on these error curves. It is obvious therefore that what Krishnan obtained is not any new effect but the result of omitting to take into consideration the error due to the convergence of light. His



statement in the paper that the error $\theta^2/2$ is only 0·03 per cent. seems to have no value, for the obvious reason that in another place in the same paper, he says that no difference in ρ_u, ρ_v, or ρ_h could be detected

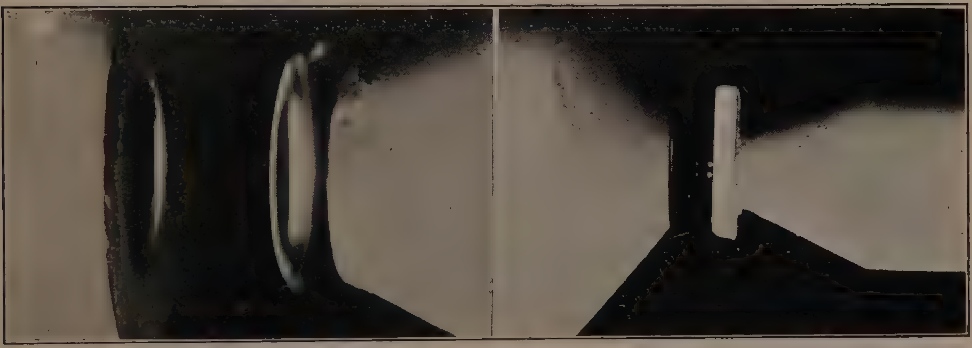
FIG. 2.



a. *b.*

Riemann mirrors in front of plane wall.

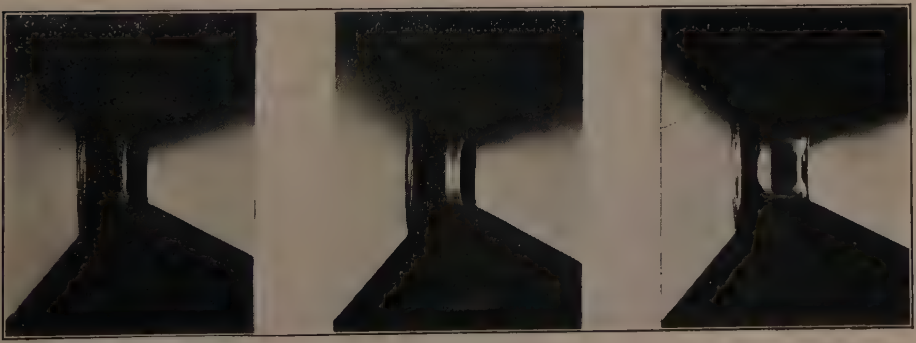
FIG. 3.



a. *b.*

Photograph of a Riemann mirror (*a*) and the reflexion of two light-points in the mirror (*b*).

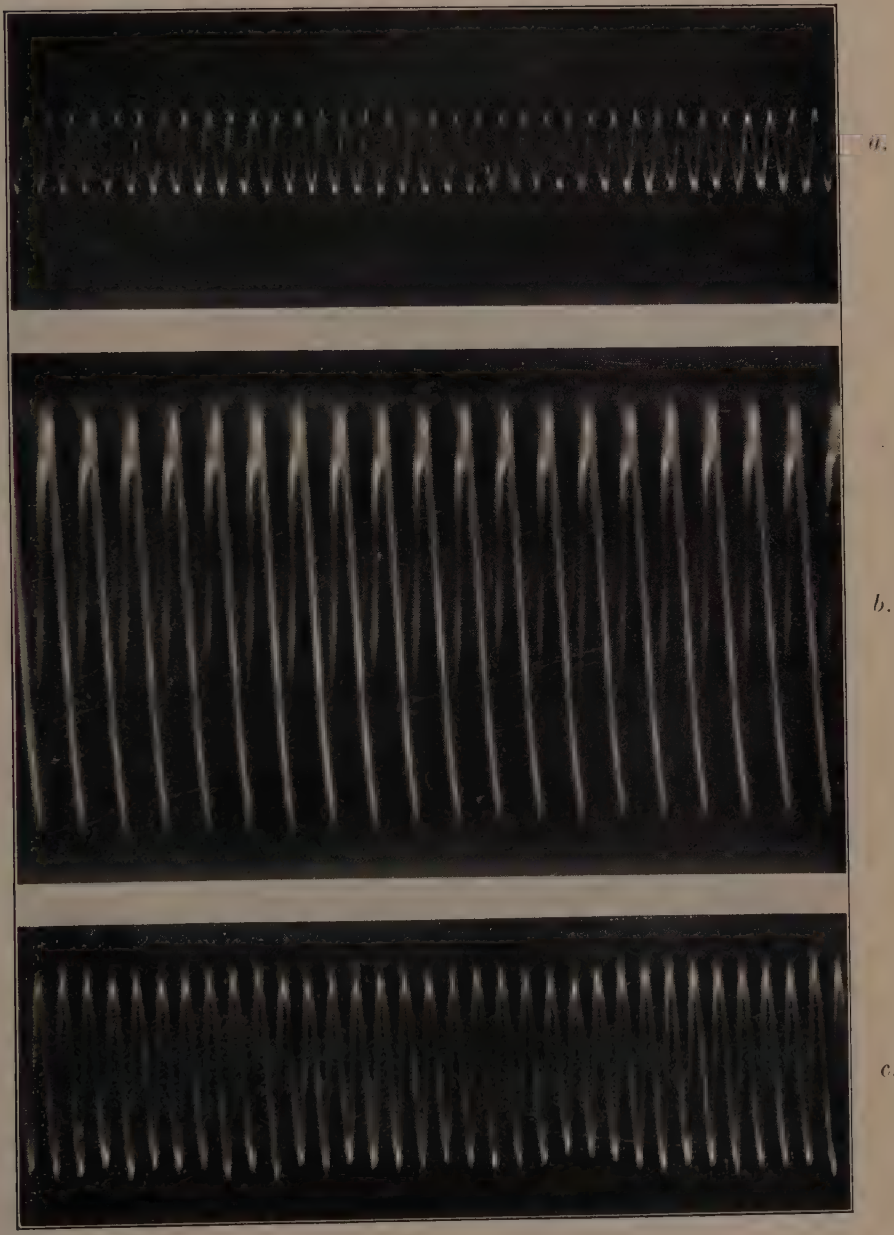
FIG. 4.



a. *b.* *c.*

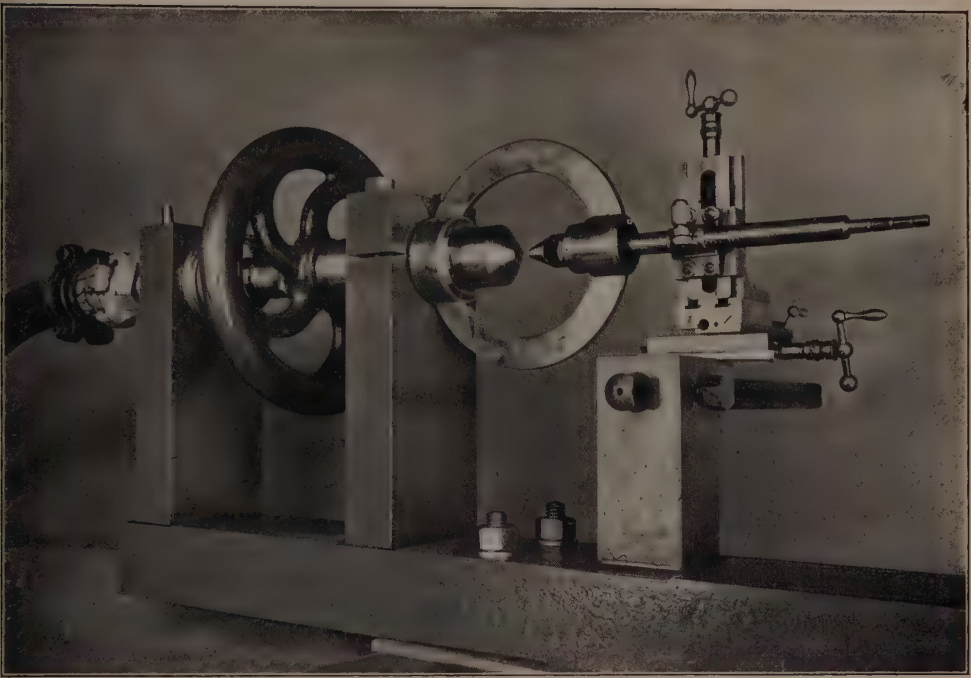
Riemann mirror in front of the oscillator of the air-jet generator.

Fig. 5.



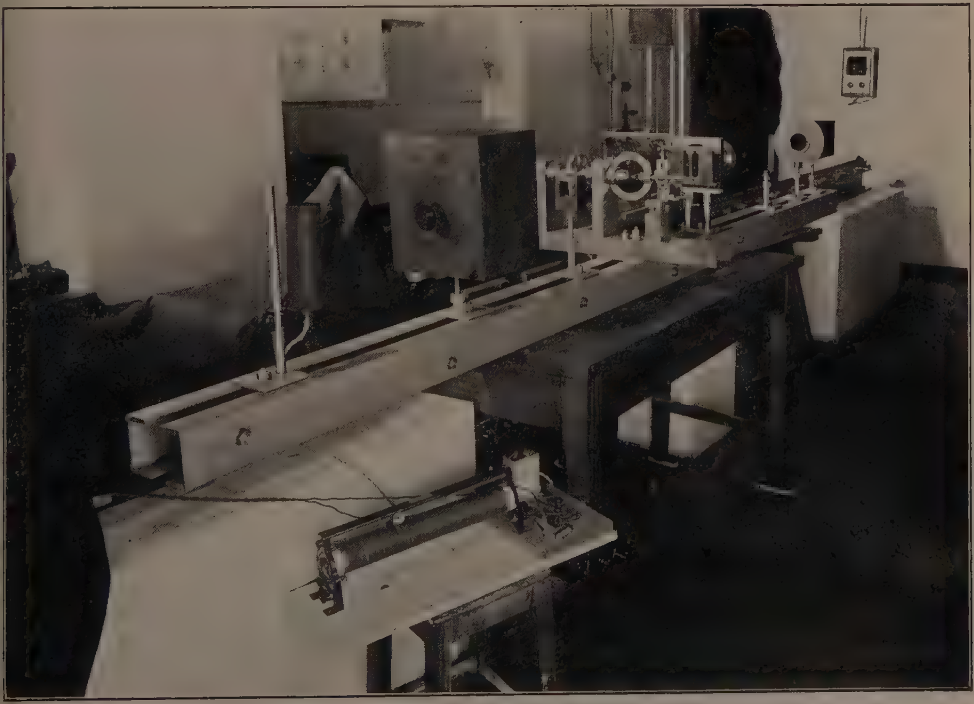
Oscillograms of the aerial vibrations in the oscillator.

FIG. 6.



Apparatus for surveying the pressure and the temperature distribution within the oscillator during operation.

FIG. 9.



Optical bench for the observation of the amplitude of the vibrations of the air in the oscillator.

even though the aperture was increased four times. In the latter case t is 0·50 per cent., and this could not have escaped his attention when he could measure 0·07 per cent. !

Regarding mixture (h), he himself says that the opalescence was very feeble ; and coupled with the fact that the temperature was lowered very near to the freezing-point of one of the components, the indications of the graph cannot be mistaken.

Further, if one looks very carefully at his results, one finds that he gets $\rho_h > 1$ in cases where ρ_v also is very small, or where there are impurities, as in tap-water, milk, and others. There is therefore no support to his contention that his method is very delicate for the detection of clusters of molecules, not small in comparison with the wave-length of light.

We do not know of any correlation *existing* between the semi-angle of convergence of the lens and size of the clusters, if one presumes to suggest that the relation between ρ_h and ρ_r would be identical in both the cases.

At best Krishnan's results are only a verification of Gans's theory regarding the necessity for correction due to convergence of the lens in all light-scattering work.

In view of the above remarks, and the explanation of Krishnan's anomalous results, we believe that no new theory or hypothesis as explained in the opening paragraph is called for.

Conclusion.

It is shown in this paper that the scattering of light observed by R. S. Krishnan in binary liquid mixtures is purely molecular, and that no scattering arises from clusters, if any, at the critical solution temperature. Neither is the critical solution mixture emulsoidal. At any rate, his optical method of using incident horizontally polarized light is vitiated by grave experimental errors which have not been taken into account. His anomalous results have been explained by taking into consideration the error due to the convergence of the lens used for focusing the incident light. No new theory therefore, as, for example, that of Frenkel or of Gans or of Mueller, appears to be necessary to explain the results.

References.

J. Frenkel, ‘ Acta Physicochimica ’ (U.R.S.S.) viii. p. 261 (1938).

R. Gans, *Phys. Zeit.* xxxvii. p. 19 (1936). See also V. S. Vrkljan and M. Katalinic, *Phys. Zeit.* xxxvii. p. 482 (1936).

R. Gans, *Phys. Zeit.* xxviii. p. 661 (1927). See also R. Ananthakrishnan, Proc. Ind. Acad. Sci. ii. p. 133 (1935).

R. S. Krishnan, Proc. Ind. Acad. Sci. i. p. 211 (1934) ; i. p. 915 (1935) ; ii. p. 221 (1935) ; and other papers of a similar kind in these Proceedings ; *Kolloid Zeit.* lxxxiv. p. 8 (1938) ; Thesis, Madras University (1937).

Hans Mueller, Proc. Roy. Soc. A, clxvi. p. 425 (1938).

XVIII. *The Musical Scale of the Highland Bagpipe.*

G. E. Allan, D.Sc., Applied Physics Department,
Glasgow University *.

[Received May 30, 1939.]

Introduction.

This instrument has in its chanter a scale of nine notes usually designated by the letters : G, A, B, C, D, E, F, G′, A′, the lower A having a frequency round about 450 vibrations per second. The bagpipe has beating reeds in the three drones, and a double reed in the chanter, so that the tones of the chanter are similar in quality to those of the oboe. Dr. A. J. Ellis, writing in 1885⁽¹⁾ , described a test which was made in conjunction with Mr. Charles Keene, the well-known 'Punch' artist, who was a bagpipe player, and Mr. A. J. Hipkins of Messrs. Broadwood, who determined the pitch of the notes by means of Dr. Ellis's set of tuning-forks. The observed frequencies and their intervals were :

	G	A	B	C	D	E	F	G′	A′
Frequency	395	441	494	537	587	662	722	790	882
Interval in cents .		191	197	145	154	208	150	156	191

On applying to Messrs. Glen of Edinburgh, they received the opinion that "If a chaunter was made perfect in any one scale, it would not go well with the drones. Also there could not be nearly so much music produced, as at present it adapts itself to the keys of A major, D major, B minor, G major, E minor and A minor." Obviously what was wanted was quantity of music and not quality. Colin Brown, of Anderson's College, Glasgow, informed Dr. Ellis that "No scientific principle was adopted in boring the holes of the chaunters, and that only about one chaunter in six turned out useful." He himself thought that the bagpipe ought to play the major scale A.

In an article on the bagpipe, in Grove's 'Dictionary of Music,' D. J. Blaikley, of Messrs. Boosey and Co., remarks that the bagpipe seems to have come into general favour in Scotland at the close of the 15th century. He notes also that Shakespeare occasionally alludes to the bagpipe and to the antipathy which some people have to its sound. A very clear account of the instrument is given. He states that the nine notes "do not form any diatonic scale whatever, nor are they accurately tuned to one another." Further, "there is good ground for believing that any attempt to accommodate the bagpipe to modern scale notation would only result in a total loss of its archaic, semi-barbarous and stimulating character." The instrument has certainly retained its pristine simple form, but no change in the scale of the chanter will take away the skirl of the pipes which is inherent in the tones of the reeds. Further, if the untunefulness of the scale in its present form were removed, so also would the irritation which it causes to musical people, whereas non-musical listeners would not notice the change. Blaikley repeats Dr. Ellis's record of Keene's chanter, and adds one which is the mean of three records of modern chanters taken by himself, namely,

---

* Communicated by the Author.

G A B C D D F G′ A′
204 193 149 157 199 148 156 198

The intervals are again expressed in cents. The two scales have this common feature, that they consist of four full-tone intervals and four intervals, each of about three-quarters of a full tone.

In a book on the Highland Bagpipe by W. L. Manson, there is an appendix on the Scale of the Pipes by J. MacNeill. He states that, in spite of attempts to modify the instrument, as has been done with the oboe, the scale remains practically the same as it was two centuries ago. He considers, however, that it does not differ so widely from the natural or equal-temperament scales as critics allege. The key probably used first was that of D, and the most convenient scale to play on is that of A. MacNeill submits a true pipe scale which, if expressed in the two notations, would run :

Staff	G	A	B	C♯	D	E	F♯	G′	A′
Sol-fa, Key D ..	f_1	s_1	l_1	t_1	d	r	m	f	s
Key A ..	ta_1	d	r	m	f	s	l	ta	d^1

The possibilities are pointed out that chanters may not always be bored with perfect accuracy, and that the holes may become enlarged by wear.

Scales of some Bagpipe Chanters.

A description is here given of frequency tests on the scales of some bagpipe chanters. The chanters tested are numbered 1, 2, etc., up to 8, and they were played, some by pipe-majors and some by amateurs (in the original sense of the word). Some were played attached to the bag in the usual way, and some without the use of the bag, *i. e.*, by mouth direct. The frequencies found were measured by the monochord(2), and the notes were played sometimes in ascending order of pitch, sometimes in the opposite direction. The simplest method of fingering was employed, the holes being covered or uncovered in the order of their position. The intervals between successive notes on the scales are again expressed in cents, as this gives a means of general comparison not possessed by a table of frequencies.

The Cent.

The cent is a very small musical interval and is such that the equal-temperament semitone contains 100 cents and the full tone 200. The octave is thus divided into 1200 cents. If, for example, we know the frequencies N_1 and N_2 of two notes, it can be shown that the musical interval between the notes, expressed in cents, is given by

$$\log \text{interval} = \log (\log N_1 - \log N_2) + 3{\cdot}6006.$$

TABLE I.—Intervals in the Scales of Eight Chanters, in Cents.

Chanter.	G	A	B	C	D	E	F	G′	A′
1...........	200	186	187	170	171	198	229	92	
2...........	203	185	176	154	184	197	104	201	
3...........	191	199	181	159	175	201	183	128	
4...........	201	205	191	147	183	173	195	120	
5...........	179	197	193	137	187	189	109	209	
6...........	210	190	189	162	179	188	179	145	
7...........	184	201	171	159	179	208	136	180	
8...........	210	183	177	149	178	201	171	132	
Keene's	191	197	145	154	208	150	156	191	
Blaikley's ..	204	193	149	157	199	148	156	198	
Equal temperament ..	200	200	200	100	200	200	100	200	

If we wish to compare the notes of the chanter with those of an equal-temperament scale, the latter will contain two semitones of 100-cent intervals and the others will be 200-cent intervals.

Intervals in Scales of Chanters.

Table I. contains the intervals as measured for the scales of the eight chanters.

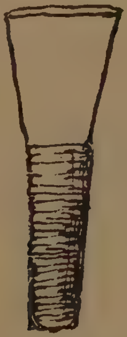
An examination of the figures in that table leads to the conclusion that no definite system was employed in producing the scales, and that there is considerable irregularity in the intervals. The scales have, however, this gratifying feature in common that the whole interval G/A′ approximates closely, in nearly every case, to the required 1400 cents.

Test for Tunefulness.

In the diatonic scale, the major tone has a frequency ratio of 9/8, or an interval of 204 cents, the minor tone 10/9 or 183 cents, and the diatonic semitone 16/15 or 112 cents. Thus we might regard as a passable full tone any interval of 183 to 204 cents, and on the same basis a semitone as one with an interval of 100 to 112 cents. Applying this test to Table I., it is found that of the intervals tested 54 per cent. do not conform to the given standard of tune, those most out of tune being the intervals C/D, F/G′ and G′/A′. The first two are more nearly three-quarter tones, and are probably made so intentionally. In nearly every case the interval G′/A′ is too small. It is worthy of remark that, of the chanters tested, No. 5 approaches quite closely to the equal temperament scale. We have been given a reason for the expanded semitones, namely, that with their aid a greater diversity of keynotes can be employed. It is somewhat otherwise with the interval G′/A′. This point will be raised later.

The Reed.

The reed consists of two thin pieces of cane or other material of somewhat triangular shape, and of the size shown in the figure. The upper part is the double vibrator, the lower ends of which are bound by strong hemp thread to a short length of copper tube which has an opening of about 4 mm. (1/6 inch) in diameter. The thread wrapping is varnished, and the lower end, on which a few turns of unvarnished thread are wound, is inserted in the "reed-seat," as the upper aperture of the pipe is called.



Double reed of chanter, full size.

Note of the Reed.

When the reed is blown by itself, it gives a shrill high-pitched note. With three reeds tested, the notes had frequencies of about 1250, 1400, and 1530 respectively. The lower octaves of these notes were also heard, and the latter are assumed to be the fundamental tones of the reeds. This appears to be a case in which the fundamental is weak, compared with its first overtone, as happens in the scales of the bugle, horn, and large bell.

The Pipe.

The pipe is made of ebony. It is about 14·3 inches long and has a conical hollow. The diameter of the hollow tube varies from $\frac{1}{4}$ inch at the upper end to $\frac{7}{8}$ inch at the lower end The pipe has a series of openings along its length like other simple wind instruments. In addition, there is a hole nearest the reed-seat and on the opposite side of the chanter from the other holes. This is the thumb-hole, and it is closed by the thumb of the left hand. There are also cross-holes near the lower end, bored at right angles to the direction of the other holes. They determine the pitch of the lowest note of the chanter.

Notes of the Pipe.

If one blows across the reed-seat, after the manner of pan-pipes, the resonant tone of the hollow tube is sounded, and a series of notes is obtained by fingering the holes as when playing the chanter. These notes vary in frequency from about 580 to about 2260.

Reed and Pipe.

After some experience with four of the eight chanters, it was found that graphs made by plotting frequency against distance of mid-hole from reed-tip were roughly coincident, following for part of their course a hyperbolic law. A new chanter was made for which the hole distances were calculated from the apparent law thus obtained. The results are given as separate points on the graph. The tuning was not considered satisfactory, and a second chanter was made, the particulars of which are given in Table II. and in the graph.

Before measuring each of the above frequencies, the reed was dried in a current of warm air and then inserted in the reed-seat. The chanter was blown by mouth direct. The closeness to the graph of the set of points ⊙ ⊙ ⊙ for the first chanter indicates the general nature of the relation of frequency to hole distance. There is a fairly regular increase of frequency as the effective length of the pipe is reduced, followed by a greater rate of increase between F and G′, and a falling off between G′ and A′. This deviation appears to be characteristic of the double reed and pipe system ; in some cases the deviation is more pronounced than in others. This effect may be the real reason why the interval G′/A′ is in most cases too small.

The Coupled System.

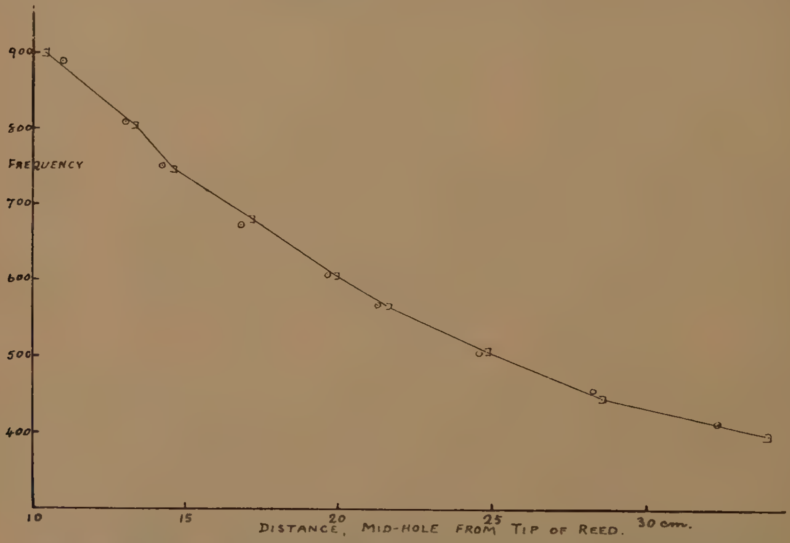
When the reed is inserted in the reed-seat and the chanter sounded, the two vibrators—double reed and pipe—form a coupled system whose units react on one another, causing a change in the pitch of the notes heard. For example, the reed with a fundamental frequency of about 700 is coupled to a pipe with a resonance frequency of 580, and the tone heard has a frequency of about 400. This last is the "coupling tone." It is due to a forced vibration and is flatter in pitch than the resonance tone of the pipe. A similar effect is present throughout the scale, but in the neighbourhood of F, where the resonance frequency of the shortening tube approaches that of the prominent second harmonic of the reed, there is, as the graph shows, a disturbance or tone displacement.

Table II.—Frequency of Chanter Note (N) and Distance of Mid-hole from Reed-tip (d).

	d cm.	Note.	N.	Interval in cents.
Reed-tip ...	0	—	—	—
Reed-seat	3·52	—	—	—
Thumb-hole ..	10·39	A	900	—
				195
Second hole ..	13·36	G	804	
				128
Third hole ...	14·60	F	747	
				186
Fourth hole ..	17·17	E	671	
				179
Fifth hole ...	19·95	D	605	
				119
Sixth hole ...	21·60	C	565	
				194
Seventh hole .	24·85	B	505	
				219
Eighth hole .	28·54	A	445	
				193
Cross-holes ...	33·90	G	398	—
Lower end ..	39·87	—	—	—

M. Wien (3) has developed the theory of coupled systems from which he has established the laws of coupling. These laws have been verified experimentally by H. Vogel (4) for the free-reed organ pipe. He also found that with passage through the resonance position there is a sudden change in frequency (a drop in pitch) as the tube is shortened. In the reed-chanter system no such drop in frequency was observed. This may be due to the double form of the reed, since the two reeds will react on one another in much the same way as the two prongs of a tuning-fork. As the reed-note is rich in harmonics, and contains the complete series of these, it is probable that the disturbance which takes place near the top of the chanter scale is due to the reaction between the first overtone of the reed and the resonance tones of the pipe. We repeat here some remarks on the free reed taken from Geiger and Scheel (5).

"The tone produced by blowing is not usually the natural tone of the free reed, nor one of the natural tones of the pipe, but that of the less damped of the two coupling tones. Its pitch is determined by the natural tone of the reed and that partial of the pipe which lies nearest to the tone of the reed. The mutual effect is not confined to the immediate neighbourhood of the position of resonance, and it gives rise to considerable tone displacement."



Frequency of reed-pipe and distance of mid-hole from tip of reed.

Intervals obtained.

The intervals which have been obtained in the above tests are, it must be admitted, not strictly in tune, although the deviations from correct intonation are smaller than those in Table I. But the figures and graph may be employed to obtain better results, in the following way:—Using the formula on p. 156 and beginning at the lower end of the table, the note G 398 is taken and a note 200 cents above G is found from : log N =

log 398+0·0502. This gives N=446·8, corresponding on the graph to a hole distance 29·00 cm. In a similar way the other hole distances can be found, using for the semitone intervals the equation $\log N_1 = \log N_2 + 0\cdot0250$. The distances so found would then be as in Table III., in which the lower line gives the distances in inches.

Table III.—Calculated Distances from Mid-hole to Reed-tip.

G	A	B	C	D	E	F	G	A
(33·90)	29·00	24·70	21·70	20·30	17·15	14·50	13·55	10·65
13·55	11·42	9·73	8·55	7·98	6·76	5·71	5·34	4·19

Some Causes of Inconstancy of Pitch.

It is well known that the thin reed is uncertain when used as a standard of pitch. Its frequency depends, among other things, on the pressure of the air current which is used to sound it, and a rise in pressure causes a rise in pitch which may amount on occasion to a semitone. The use of the bag will, no doubt, make for greater steadiness of pitch. Moisture on the reed, from the breath, lowers the pitch of the scale, and it is usual among pipers to season or mature a new reed to bring it into a constant condition. Further, a change of reed may cause a change of pitch.

Conclusion.

The chanter which has resulted from these tests is one in which the more accurate tuning of the semitones disturbs the nearly regular spacing of the finger-holes. This renders the fingering more difficult, and possibly supplies the reason for preferring three-quarter tones to semitones. The bagpipe maker has, therefore, several alternatives in making a chanter. He may, as before, make a simple, easily fingered chanter which is not quite in tune. Or he may make a chanter on the above lines which is in tune but more difficult to finger. A third method would be to attain equalization of the finger spacing by the use of metal levered stops as in the oboe, and a fourth method would be to have the chanter hollowed in such a way as to attain equal spacing.

I acknowledge with pleasure the help received, in making the tests, from Mr. R. G. Lawrie, Pipe-Major MacDougall and Mr. D. Johnston, of Glasgow.

Literature.

(1) A. J. Ellis, Roy. Soc. Arts J. pp. 485–527 (1885).
(2) Phil. Mag., Dec. 1927 and Oct. 1938.
(3) M. Wien, *Ann. d. Physik u. Chem. N.F.* lxi. pp. 151–189 (1897).
(4) H. Vogel, *Ann. d. Physik u. Chem. N.F.* lxii. pp. 247–282 (1920).
(5) Geiger and Scheel, *Handbuch der Physik*, viii. p. 283.

XIX. *A Note on the Carey-Foster Method of Measuring a Mutual Inductance in Terms of a Capacity.*

By E. Tyler, D.Sc., M.Sc., F.Inst.P., A.F.R.Ae.S.,
Science Department, South-East Essex Technical College, Dagenham *.

[Received June 10, 1939.]

Despite the recent criticism † directed against the Carey-Foster method for comparison of a mutual inductance M in terms of a capacity C, the method merits a place among the list of recognized academic experiments.

If we dispense with the usual procedure of substituting the experimental observations in the formula relating M and C and adopt graphical means of analysing the results, the method becomes more interesting and instructive.

This note pertains to suggested graphical methods whereby coordination of the experimental observations permits the following :—

1. Computation of M in terms of C without a knowledge of the resistance of the secondary winding of the inductance.
2. Deduction of the resistance of the secondary winding without actual measurement.
3. Comparison of capacities.

Fig. 1 is a typical Carey-Foster circuit in which M is the mutual inductance and C the capacity under comparison.

The normal procedure is to adjust the resistances R_1 and R_2 until there is no deflexion of the galvanometer G at either the make, break, or reversal of the primary current. Under such conditions of balance it is readily shown that

$$M=C\,.\,R_1(R_2+S)\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and substituting for C, R_1, R_2, and S suffices to obtain M.

A separate experiment is thus essential for the evaluation of S, and it is only when R_2 is large compared with S can we possibly ignore the latter.

Method 1.

Now rearranging equation (1) we may write it in the form

$$R_2+S=\frac{M}{C}\cdot\frac{1}{R_1},\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

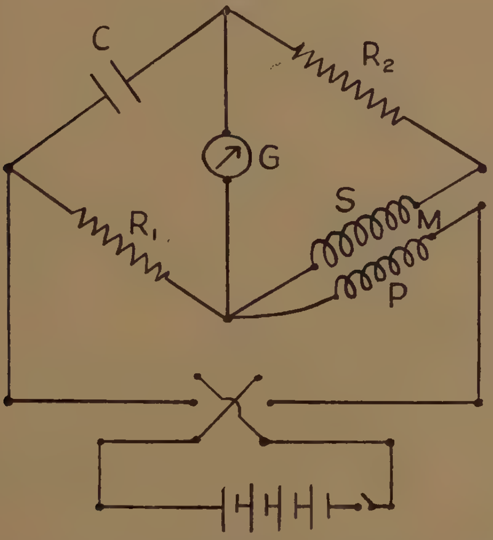
---

* Communicated by the Author.

† Astbury and Ford, Journ. Sci. Inst. vol. xv. p. 94, no. 3 (March 1938).

and this equation is linear in R_2 and $\frac{1}{R_1}$, since S and $\frac{M}{C}$ are constant. Hence, if a series of values of R_1 and R_2 are obtained for the balanced conditions, and a plot of R_2 against $\frac{1}{R_1}$ is made, the resulting graph should be a straight line not passing through the origin (fig. 2). The slope of this graph is $\frac{M}{C}$ and is independent of S. The intercept corresponding to $\frac{1}{R_1}=0$ gives a value of S, and, provided S is comparable with R_2, this affords a reliable method of deriving S, substitution of which can afterwards be made in equation (1), if desired. It may not, however, be always convenient to realise this condition.

Fig. 1.

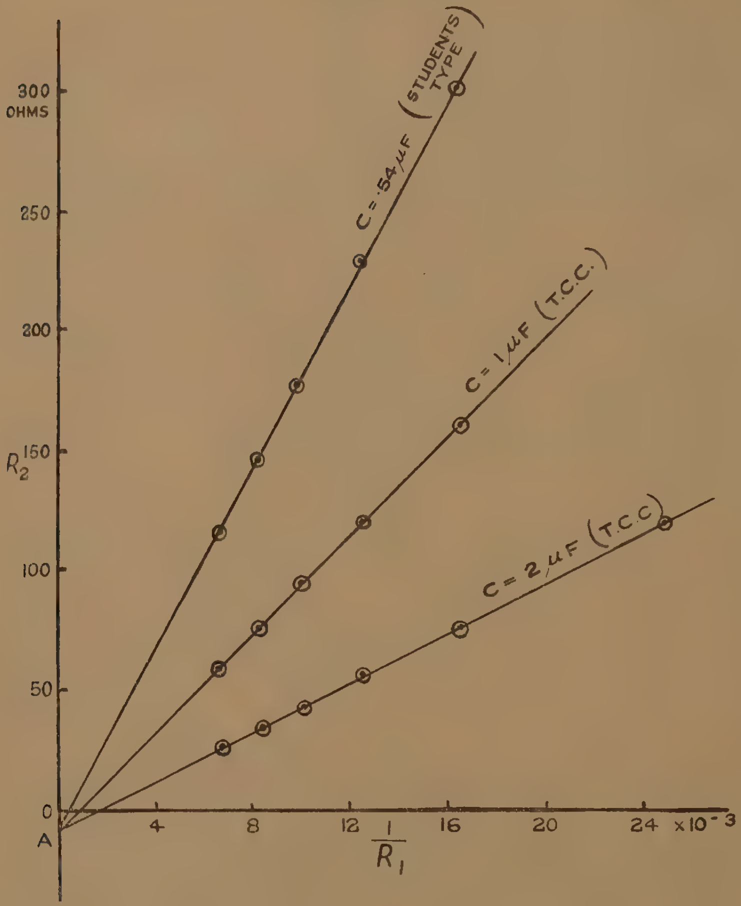


By repeating the experiment with other capacities we obtain a family of straight lines of different slopes but all intersecting at the same point on the R_2 axis corresponding to $\frac{1}{R_1}=0$, *i. e.*, giving the same intercept OA equivalent to S (fig. 2).

Furthermore, comparison of capacities is readily effected, since, denoting any pair of capacities by the suffixes, and remembering M=C×slope, which is constant, we derive the condition

$$\frac{C_1}{C_2}=\frac{(\text{slope})_2}{(\text{slope})_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Fig. 2.



Method 2.

Equation (1) may also be re-written in the form

$$R_1R_2=\frac{M}{C}-R_1S, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

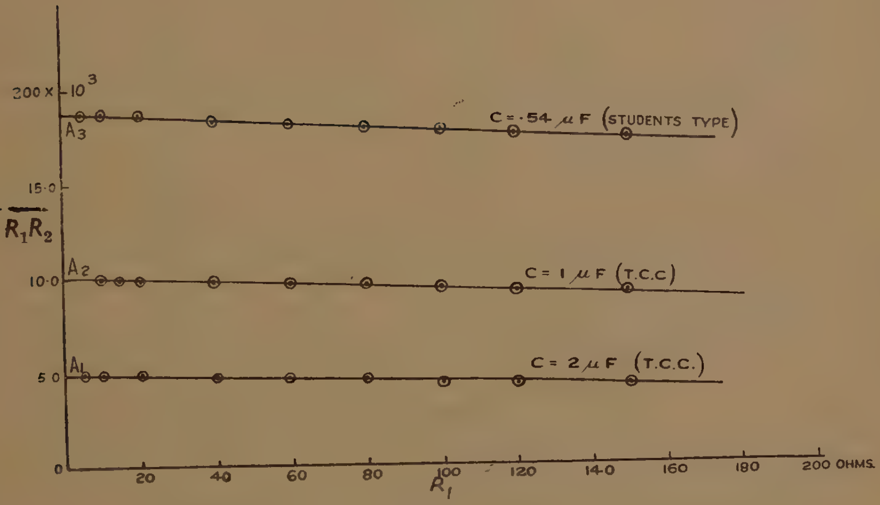
and this represents a linear relationship between R_1R_2 and R_1, since $\frac{M}{C}$ and S remain constant. A plot of R_1R_2 against R_1 should therefore produce a straight line graph with negative slope S and intercept $\frac{M}{C}$ on the R_1R_2 axis corresponding to $R_1=0$ (fig. 3). The intercept on the R_1 axis corresponding to $R_1R_2=0$ is a measure of $\frac{M}{C}\cdot\frac{1}{S}$.

The derivation of both S and $\frac{M}{C}$ is therefore direct and this method is to be preferred, particularly for the evaluation of S.

With other capacities $C_1 C_2$ etc., and same M, a family of parallel straight lines are obtained, all of slope S and vertical intercepts OA_1, OA_2 etc. of magnitudes $\frac{M}{C_1}$, $\frac{M}{C_1}$ etc., whence

$$\frac{C_1}{C_2} = \frac{OA_2}{OA_1} \quad \ldots \ldots \ldots \ldots \quad (5)$$

Fig. 3.



METHOD 3.

Equation (2) may also be recast as

$$\frac{1}{R_2} = \frac{M}{SC} \cdot \frac{1}{R_1 R_2} - \frac{1}{S} \quad \ldots \ldots \ldots \quad (6)$$

upon dividing throughout by $R_2 S$.

The variables of this equation are $\frac{1}{R_2}$ and $\frac{1}{R_1 R_2}$, hence plotting these quantities gives a straight line graph (fig. 4).

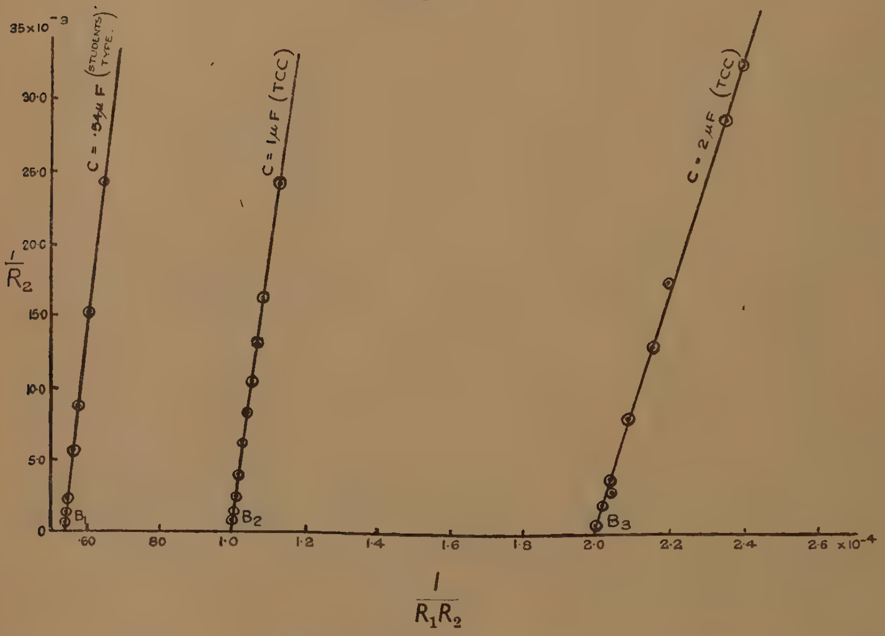
We cannot obtain $\frac{M}{C}$ direct from the slope since this is $\frac{M}{SC}$, and S must be known. The intercept of $\frac{1}{R_2}$ corresponding to $\frac{1}{R_1 R_2} = 0$ does, however, give $\frac{1}{S}$, and that of $\frac{1}{OB}$ corresponding to $\frac{1}{R_2} = 0$ gives $\frac{M}{C}$.

With different capacities we obtain a series of non-parallel straight lines all intersecting at the same point on the $\frac{1}{R_2}$ axis corresponding to $\frac{1}{S}$.

For comparison purposes we employ the intercept condition $\frac{M}{C}=\frac{1}{OB}$, which for two capacities, C_1 and C_2, reduces to

$$\frac{C_1}{C_2}=\frac{OB_1}{OB_2} \quad \ldots\ldots\ldots\ldots \quad (7)$$

Fig. 4.



Specimen Results.

These are correlated in Tables I. & II. and figs. 2–4.

The deduced values of M in terms of C from the slopes of the graphs (fig. 2) compare well with those obtained by direct substitution in equation (1).

Despite S being small compared with R_2, it was possible to deduce S from the intercept by only plotting the lower values of R_1 against $\frac{1}{R_2}$ using as large a scale as possible. The interpolated value of S (=8·0 ohms) agrees well with the PO box determination.

In fig. 3 the vertical intercepts OA_1, OA_2, etc., and the slopes, each

TABLE I.

C=·54 μF. (Students type, Philip Harris.) S=8·0 ohms. (PO box).			C=IμF. (T.C.C.) S=8·0 ohms.			C=2μF. (T.C.C.) S=8·0 ohms.		
R_1.	R_2.	$M=CR_1(R_2+S)$.	R_1.	R_2.	$M=CR_1(R_2+S)$.	R_1.	R_2.	$M=CR_1(R_2+S)$.
ohms. 5	ohms. 3720	milli-henries. 10·06	ohms. 5	ohms. 2000	milli-henries. 10·04	ohms. 5	ohms. 990	milli-henries. 9·98
10	1865	10·11	10	1000	10·08	10	495	10·06
15	1240	10·10	15	660	10·02	15	325	9·99
20	925	10·07	20	495	10·06	20	245	10·12
40	455	10·00	40	245	10·12	40	120	10·24
60	300	9·97	60	163	10·26	60	77	10·20
80	223	9·97	80	120	10·24	80	57	10·40
100	177	9·99	100	95	10·30	100	42	10·00
120	146	9·97	120	77	10·20	120	35	10·32
150	116	10·04	150	61	10·35	150	27	10·50
* Mean=10·02.			Mean=10·16.			Mean=10·18.		

* Inductance used was labelled 10 milli-henries (Griffin and Tatlock).

equivalent to $\frac{M}{C_1}$, $\frac{M}{C_2}$, and S respectively, are also concordant with the values obtained by method 1.

TABLE II.

Fig.	Capacity C.	Slope$=\frac{M}{C}$.	M=C×slope.	S=OA.
	μF.		milli-henries.	ohms.
2	·54	$1{\cdot}85\times10^4$	10·04	8·0
	1·00	$1{\cdot}02\times10^4$	10·20	8·0
	2·00	$\cdot54\times10^4$	10·20	8·0
		Intercept OA$=\frac{M}{C}$.	M=C×OA.	Slope=S.
			milli-henries.	ohms.
3	·54	$1{\cdot}87\times10^4$	10·09	8·40
	1·00	$1{\cdot}01\times10^4$	10·10	8·00
	2·00	$\cdot54\times10^4$	10·00	7·50
		Intercept OB$=\frac{C}{M}$.	M$=\frac{C}{OB}$.	S.
			milli-henries	ohms.
4	·54	$\cdot535\times10^{-4}$	10·09	
	1·00	$1{\cdot}00 \quad 10^{-4}$	10·00	unreliable
	2·00	$1{\cdot}99 \quad 10^{-4}$	10·05	

Method 3 does not yield so reliable results from the slopes mainly because of the relative low value of S and the small range of R_1R_2. The smaller the capacity the bigger the slope of the graph for a given range of $\frac{1}{R_2}$, *i. e.*, smaller variation of R_1R_2, and possible larger error in the computed intercept corresponding to $\frac{1}{S}$. Furthermore, since the slope incorporates S as well as $\frac{M}{C}$, we must regard methods 1 and 2 as the more reliable for obtaining $\frac{M}{C}$ from this point of view. For method 3 to be successful in this respect S must be large and comparable with R_2, R_1R_2 must be variable over a wide range, and the comparison capacity also fairly large. Reliability of $\frac{M}{C}$ from the reciprocal of the intercept OB is, however, satisfactory.

XX. *Interaction between Radiation and Matter.*

By Prof. K. C. Kar, D.Sc.[*]

[Received February 13, 1939.]

Quantum mechanical theories of interaction between radiation and matter have been developed by Dirac [†], Heisenberg [‡], Fermi [§], Jordon and Pauli [‖], and others. Some of these problems will be discussed in the present paper from the point of view of wave-statistics. Before going into the problems it will be necessary to explain briefly the principles of the wave-statistics, already discussed in some of my earlier papers [¶].

1. *Phase Space for a Group.*

It is well known in statistical mechanics that the $\{\mu\}$-space for a group of particles (molecules, atoms, etc.) is considered as a hydrodynamical fluid with a six-dimensional equation of continuity. Thus, as in hydrodynamics, we have

$$\frac{\partial D}{\partial t}+\sum_s^{1,3}\left\{\frac{\partial}{\partial q_s}(\dot{q}_s D)+\frac{\partial}{\partial p_s}(\dot{p}_s D)\right\}=0, \quad . \quad . \quad . \quad . \quad (1)$$

where D is the phase density and q_s, p_s are the positional and momenta coordinates. The above may also be written as

$$\frac{D}{Dt}(D)+D\sum_s^{1,3}\left(\frac{\partial \dot{q}_s}{\partial q_s}+\frac{\partial \dot{p}_s}{\partial p_s}\right)=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where

$$\frac{D}{Dt}=\frac{\partial}{\partial t}+\sum_s^{1,3}\left(\dot{q}_s\frac{\partial}{\partial q_s}+\dot{p}_s\frac{\partial}{\partial p_s}\right).$$

---

[*] Communicated by the Author

[†] Dirac, Proc. Roy. Soc. A, cxiv. p. 243 (1927) *et seq.*

[‡] W. Heisenberg, *Ann. d. Phys.* ix. p. 338 (1931); *Zeits. f. Phys.* lxv. p. 4 (1930).

[§] E. Fermi, *Lincei Rend.* x. p. 72 (1929); ix. p. 881 (1929); xii. p. 431 (1930).

[‖] Jordon and Pauli, *Zeits. f. Phys.* xlvii. p. 151 (1928).

[¶] K. C. Kar, *Zeits. f. Phys.* lix. p. 102 (1929); lx. p. 243 (1930); lxi. p. 675 (1930); lxii. p. 855 (1930); Kar and Mukherjee, Phil. Mag. xvii. p. 993 (1934); Kar, Ind. Phys. Math. Journ. iii. p. 65 (1932).

Liouville has taken the phase point to be a conservative system. Accordingly the second term in (2) vanishes and we have

$$\frac{D}{Dt}(D)=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

being the well-known condition of incompressibility. Now, although the total energy of the assembly is conserved, we are not justified in assuming that the energy in an elementary phase volume is constant. This is obvious when we remember that spontaneous fluctuation within the assembly is a fundamental characteristic of statistics. We should, therefore, take the Hamiltonian for a non-conservative system, viz.,

$$\left.\begin{aligned} \dot{p}_s &= -\frac{\partial E}{\partial q_s}-\frac{\partial F}{\partial \dot{q}_s}, \\ \dot{q}_s &= \frac{\partial E}{\partial p_s}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where E is the energy and F the dissipation-function. Hence from (2) and (4) we have, in place of Liouville's Eq. (3),

$$\frac{D}{Dt}(D)=D\sum_s^{1,3}\frac{d^2F}{\partial p_s \partial \dot{q}_s}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

showing that the phase fluid is compressible. It may be pointed out that the q-portion of the $\{\mu\}$-space possesses all the properties of the matter it represents. By analogy we may attribute similar properties also to the p-space. Thus it appears natural that the phase fluid representing matter should possess all its properties. There is, therefore, no difficulty in imagining a compressible phase space. Now we have from (5)

$$\frac{\partial D}{\partial t}+\sum_s^{1,3}\left(\frac{\partial E}{\partial p_s}\cdot\frac{\partial D}{\partial q_s}-\frac{\partial E}{\partial q_s}\cdot\frac{\partial D}{\partial p_s}\right)=D\sum_s^{1,3}\frac{d^2F}{\partial p_s \partial \dot{q}_s}+\sum_s^{1,3}\frac{\partial D}{\partial p_s}\frac{\partial F}{\partial \dot{q}_s}. \quad (5.1)$$

As the second term in the right-hand side is negligible compared with the first, the above becomes

$$\frac{\partial D}{\partial}+\sum_s^{1,3}\left(\frac{\partial E}{\partial p_s}\cdot\frac{\partial D}{\partial q_s}-\frac{\partial E}{\partial q_s}\frac{\partial D}{\partial p_s}\right)=D\sum_s^{1,3}\frac{d^2F}{\partial p_s\cdot\partial \dot{q}_s}. \quad . \quad . \quad (5.2)$$

Consider a group of hydrogen atoms in statistical equilibrium. Here the nuclei move randomly, while the orbital electrons execute periodic motion. Because of this, and also, probably, due to the fact that the difference in the masses is very great, the motion of one group is practically unaffected by the other. Thus we can imagine two phase spaces ; one for the nucleus or, rather, the atom as a whole, and the other for the orbital electron. In the former space, owing to random motion, the average values of the dissipation-function F, and also of $\frac{\partial D}{\partial t}$, must vanish in a given elementary phase volume. We have then from (5.2)

$$\sum_{s}^{1,3}\left(\frac{\partial E}{\partial p_s}\cdot\frac{\partial D}{\partial q_s}-\frac{\partial E}{\partial q_s}\cdot\frac{\partial D}{\partial p_s}\right)=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

leading to Gibbs's canonical distribution, well known in classical statistics.

In the second space, on account of the periodic motion the phase density changes periodically, so that the dissipation-function F cannot be taken either positive or negative. It should be taken imaginary. Thus we may put $2\pi i\nu$ for $\sum_{s}^{1,3}\frac{d^2F}{\partial p_s\,\partial\dot{q}_s}$. We have then from (5.2)

$$\frac{\partial D}{\partial t}+\sum_{s}^{1,3}\left(\frac{\partial E}{\partial p_s}\cdot\frac{\partial D}{\partial q_s}-\frac{\partial E}{\partial q_s}\,\frac{\partial D}{\partial p_s}\right)=2\pi i\nu.D, \quad . \quad . \quad . \quad . \quad (7)$$

where D denotes the total phase density. It is evident that because the phase space is slightly compressible, $D=D_1D_2$, where D_1 is the steady value of the density, being the solution of (6), and D_2 the periodic component, being the solution of

$$\frac{\partial D_2}{\partial t}=2\pi i\nu.D_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

that is,

$$D_2=\chi_1\chi_2e^{2\pi i\nu t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8.1)$$

where χ_1, χ_2 are functions of q's, p's respectively. Now it has been shown by the writer * that Eq. (6) yields Fermi or Bose-Einstein values of phase density on using the quantum statistical idea of phase cell. Thus the total phase density is given by

$$D=\frac{N}{\Phi}\cdot\frac{1}{e^{\frac{\psi-E}{kT}}\pm 1}\cdot\chi_1\chi_2e^{2\pi i\nu t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where, according to Planck, the total phase volume $\Phi=Nh^3$. Thus the total average phase density, viz. $\frac{N}{\Phi}$, is constant. The second factor in (9) represents the variation of phase density with the energy, while the factors χ_1 and χ_2 give the variation within an energy shell E and E+dE in the phase space. Such a fine variation is completely ignored in quantum statistics.

From consideration of dimensions we find from (9) that χ_1, χ_2 are dimensionless. Hence we must have

$$\chi_1\left(\frac{q}{q_0}\right),\quad \chi_2\left(\frac{p}{p_0}\right),$$

---

* K. C. Kar and R. C. Mazumdar, *Zeits. f. Phys.* lv. p. 546 (1929).

and it will be shown later on that $\chi_1=\chi_2$. Now as the phase fluid is of finite volume ($\Phi=Nh^3$), the waves set up in it are reflected from the boundary and form stationary waves. Thus any phase integral must have the same value at all times. We may therefore drop the time factor when taking the average value of $\chi_1\,\chi_2$. Thus we have from (9)

$$D\Delta\tau=\frac{N}{\Phi}\,\frac{\Delta\tau}{e^{\frac{\psi-E}{kT}}\pm1}\cdot\frac{1}{\Delta\tau}\int^{\Delta\tau}\chi_1\chi_2\,d\tau, \qquad \ldots\ldots \qquad (9.1)$$

where $\Delta\tau=\Delta\tau_q.\Delta\tau_p$. As, however, $\Delta\tau_p=d\tau_p$ and χ_1, χ_2 are constant over this range, (9.1) may be written as

$$D\Delta\tau=\frac{N}{\Phi}\cdot\frac{\Delta\tau}{e^{\frac{\psi-E}{kT}}\pm1}\cdot\frac{1}{\Delta\tau_q}\int\chi_1\chi_2\,d\tau_q. \qquad \ldots\ldots \qquad (9.2)$$

It is interesting to note that (9.2) gives the quantum statistical value of phase density only when the average

$$\frac{1}{\Delta\tau_q}\int\chi_1\chi_2\,d\tau_q=1. \qquad \ldots\ldots\ldots\ldots \qquad (10)$$

So this average (normalization) condition must always be satisfied by the wave-functions χ_1, χ_2.

2. *Waves in the Phase Space.*

The waves set up in q- and p-spaces should have, in general, different frequencies. We therefore assume that in (9) $\nu=\nu_1+\nu_2$, where ν_1 and ν_2 are frequencies in the q- and p-spaces respectively. From Ehrenfest's adiabatic hypothesis, or from the theory of intermittent action [*], we find that the frequency in a stationary state is

$$\nu_1=\frac{2E_{\text{kin}}}{h}. \qquad \ldots\ldots\ldots\ldots \qquad (11)$$

Waves in the p-space are set up only when there is transition from one energy state to another. So according to Bohr $\nu_2=\frac{E_1-E_2}{h}$. Therefore, for a given energy state,

$$\nu_2=\frac{E}{h}. \qquad \ldots\ldots\ldots\ldots \qquad (12)$$

The velocity (v_1) of the χ_1-wave in q-space is evidently

$$v_1\equiv\dot{q}=\sqrt{\frac{2(E-V)}{m}} \qquad \ldots\ldots\ldots\ldots \qquad (13)$$

---

[*] K. C. Kar, *Zeits. f. Phys.* lvii. p. 416 (1929).

By analogy the velocity of the χ_2-wave in p-space is

$$v_2 \equiv \dot{p} = \frac{d}{dt}(\sqrt{2m(\mathrm{E}-\mathrm{V})}) = \frac{m}{\sqrt{2m(\mathrm{E}-\mathrm{V})}} \cdot \frac{d\mathrm{E}}{dt} . \quad . \quad . \quad . \quad (14)$$

Now the transitions contemplated take place suddenly, and may be represented mathematically by

$$\left.\begin{array}{l} \mathrm{E}=\mathrm{E}_0 e^{\pm bt}, \\ q=q_0 e^{\pm bt}, \\ . \quad . \quad . \quad . \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

where, it is well known, the reciprocal of b gives the life of the stationary state. From (15) we have, on differentiating,

$$\frac{\partial \mathrm{E}}{\partial t} = b\mathrm{E}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.1)$$

$$p = mbq. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15.2)$$

On using (15.1) we have

$$v_2 = \frac{mb\mathrm{E}}{\sqrt{2m(\mathrm{E}-\mathrm{V})}} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

Because observations are always made in q-space, we project the χ_2-wave in the q-space. The velocity of this projected wave is evidently (*vide* (15.2))

$$v_2 = \frac{\mathrm{E}}{\sqrt{2m(\mathrm{E}-\mathrm{V})}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16.1)$$

which is the phase velocity of Schrödinger. The well-known differential equation for the compressional wave in the phase space is

$$\Delta \mathrm{D} = \frac{1}{v^2} \cdot \frac{d^2\mathrm{D}}{dt^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where D is the phase density. Using the frequency and velocity given in (11) and (13), we have for the χ_1-wave, after eliminating time,

$$\Delta_q \chi_1 + \frac{8\pi^2 m}{h^2}(\mathrm{E}-\mathrm{V})\chi_1 = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

Again using the values given in (12) and (16), we have for the χ_2-wave in p-space

$$\Delta_p \chi_2 + \frac{8\pi^2 m}{(mb)^2 h^2}(\mathrm{E}-\mathrm{V})\chi_2 = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

and for the χ_2-wave in q-space

$$\Delta_q \chi_2 + \frac{8\pi^2 m}{h^2}(\mathrm{E}-\mathrm{V})\chi_2 = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (19.1)$$

Thus χ_1- and χ_2-functions are identical *. Furthermore, it may be easily verified on using the relation (15.2) that (9) is a solution of the differential equation (7).

It is clear that the χ_1- and χ_2-waves filling up the phase space are generated by the electrons themselves. Therefore the total energy of the group of electrons considered as particles must be equal to the energy of the waves integrated over the whole of the phase space. In other words, the electrons may be considered either as particles or as waves. It should of course be noted that a particle, however small it might be, when singled out of others, cannot be treated as a wave. It may be replaced by a wave only when it is found in a group in statistical equilibrium. The wave is thus essentially statistical in nature.

3. *Viscosity of the Phase Space.*

As has already been pointed out, the χ_2-wave is associated with transition from one stationary state, say mth, to the nth. So the p-component of the phase density for the mth state decreases, while that for the nth state increases correspondingly. In such a case, therefore, we are not justified in taking the dissipation function as wholly imaginary. It should be partly imaginary and partly real. Thus we have in place of (7)

$$\frac{\partial \mathrm{D}}{\partial t}+\sum_{s}^{1,3}\left(\frac{\partial \mathrm{E}}{\partial p_s}\cdot\frac{\partial \mathrm{D}}{\partial q_s}-\frac{\partial \mathrm{E}}{\partial q_s}\cdot\frac{\partial \mathrm{D}}{\partial p_s}\right)=(2\pi i\nu\mp b)\mathrm{D}. \quad . \quad . \quad . \quad (20)$$

On solving we get corresponding to (9), taking b to be negative,

$$\mathrm{D}=\mathrm{D}_0\chi_1\chi_2 e^{2\pi i\nu t-bt}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where $\mathrm{D}_0=\frac{\mathrm{N}}{\Phi}\,\frac{1}{e^{\frac{\psi-\mathrm{E}}{k\mathrm{T}}}\pm 1}$ and b has evidently the same meaning as in (15).

Formula (21) shows that the phase density for the mth state rapidly decreases. The corresponding χ_2-wave is damped as in viscous fluid. Accordingly the wave-equation is modified as in hydrodynamics, and we have, instead of (17),

$$\frac{d^2\mathrm{D}}{dt^2}+2b\frac{d\mathrm{D}}{dt}=v^2.\Delta\mathrm{D}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

where b is evidently of the nature of the viscosity coefficient. On substituting (21) in (22) and using the values of ν and v given in (12) and (16.1), we have for the χ_2-equation

$$\Delta\chi_2+\frac{8\pi^2 m}{h^2}\left\{(\mathrm{E}-\mathrm{V})+\frac{b^2h^2}{4\pi^2\mathrm{E}^2}(\mathrm{E}-\mathrm{V})\right\}\chi_2=0, \quad . \quad . \quad (23)$$

---

* The same is true for the relativistic case (*vide* Kar and Mukherjee, Phil. Mag. *loc. cit.*).

or approximately

$$\Delta\chi_2+\frac{8\pi^2m}{h^2}\left(\mathrm{E}-\mathrm{V}+\frac{b^2h^2}{4\pi^2\mathrm{E}}\right)\chi_2=0. \quad . \quad . \quad . \quad . \quad (23.1)$$

It is evident that when an electron just begins to emit or absorb radiation the p-component of the phase fluid near about, temporarily develops viscosity. And at that time the wave-function χ_2 should be taken damped as in (23.1). Because it involves the square of the viscosity coefficient, (23.1) should be same for emission $(+b)$ as well as absorption $(-b)$. It will, however, be shown later on that an elementary p-phase volume becomes viscous only when the average electromagnetic coupling potential (U) vanishes within the volume, but has a finite value outside it. It appears that this is the condition which governs emission and absorpion of radiation. It will, however, be found in §8 that in case of scattering of radiation both the p- and q-components of the phase fluid become viscous.

We are not concerned at present with the case in which only the q-component is viscous. This occurs when the phase density is very high, as in nuclei of heavier elements. It may be mentioned that on this assumption the writer * has already developed the theory of spontaneous and artificial disintegration of atom nuclei.

4. *Electromagnetic Potential.*

The electromagnetic field of a plane stationary wave has a vector potential of the form

$$\mathrm{A}=\sum_s\mathrm{A}_s=\sum_s\alpha_s u_s(t)\sin\gamma_s\,; \quad \gamma_s=\frac{2\pi\beta_s\nu_s}{c}r+\delta_s, \quad . \quad . \quad . \quad (24)$$

where α_s and β_s are respectively unit vectors in the direction of the electric force and the wave propagation. They are mutually perpendicular, since the wave is transverse. The factor $u_s(t)$ gives the dependence on time and ν_s, δ_s are the frequency and phase.

Now the electric and magnetic forces derived from the potential A are well known to be

$$\mathrm{E}=-\frac{1}{c}\frac{\partial\mathrm{A}}{\partial t}\,; \quad \mathrm{H}=\mathrm{rot}\,\mathrm{A}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

On substituting (24) in (25) we have

$$\mathrm{E}=-\sum_s\frac{1}{c}\alpha_s\dot{u}_s\sin\gamma_s,$$

$$\mathrm{H}=\sum_s\frac{2\pi\nu_s}{e}[\alpha_s,\beta_s]\cos\gamma_s.$$

---

* K. C. Kar and A. Ganguli, Phil. Mag. xvi. p. 1097 (1933) ; Kar, Phil. Mag. xxi. p. 1067 (1936) ; U. C. Bardhan, Phil. Mag. xxv. p. 1033 (1938).

Or we have, since

$$\overline{\sin^2\gamma_s}=\overline{\cos^2\gamma_s}=\tfrac{1}{2} \quad \text{and} \quad [\alpha_s, \beta_s]^2=1,$$

$$\overline{\mathrm{E}^2}=\frac{1}{2c^2}\sum_s \dot{u}_s^2 ; \quad \overline{\mathrm{H}^2}=\sum_s\frac{2\pi^2\nu_s^2}{c^2}\,.\,u_s^2.$$

Thus the electromagnetic energy in the volume Ω is

$$\mathrm{W}_1=\Omega\frac{\overline{\mathrm{E}^2}+\overline{\mathrm{H}^2}}{8\pi}=\frac{\Omega}{8\pi c^2}\sum_s(\tfrac{1}{2}\dot{u}_s^2+2\pi^2\nu_s^2u_s^2). \quad . \quad . \quad . \quad . \quad (26)$$

It is well known that the variable (v_s) conjugate to u_s is

$$v_s=\frac{\partial \mathrm{W}_1}{\partial \dot{u}_s}=\frac{\Omega}{8\pi c^2}\,.\,\dot{u}_s. \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.1)$$

Writing (26) in terms of the conjugate variables u_s, v_s, we have

$$\mathrm{W}_1=\sum_s\left(\frac{8\pi c^2}{\Omega}\,.\,\frac{v_s^2}{2}+\frac{\Omega}{8\pi c^2}\,.\,2\pi^2\nu_s^2u_s^2\right). \quad . \quad . \quad . \quad . \quad (26.2)$$

Introducing two other new conjugate variables Q_s, P_s, given by

$$u_s=\left(\frac{8\pi c^2}{\Omega}\right)^{\frac{1}{2}}\mathrm{Q}_s ; \quad v_s=\left(\frac{\Omega}{8\pi c^2}\right)^{\frac{1}{2}}.\,\mathrm{P}_s, \quad . \quad . \quad . \quad . \quad (26.3)$$

we have from (26.2)

$$\mathrm{W}_1=\sum_s(\tfrac{1}{2}\mathrm{P}_s^2+2\pi^2\nu_s^2\mathrm{Q}_s^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad (26.4)$$

On taking Q_s's, P_s's to be the positional and momenta coordinates one finds that (26.4) gives the energy of a system of oscillators of unit mass and frequencies $\nu_1, \nu_2, \ldots \nu_s$. Thus expressed in the coordinates Q_s's and P_s's the electromagnetic energy of radiation becomes equal to the energy of a system of unit oscillators of corresponding frequencies. So in building the phase space for atoms and radiation we may suppose the assembly to be a mixture of atoms and unit oscillators. Now the vector potential (24) in terms of the variables Q_s's takes the form

$$\mathrm{A}=c\left(\frac{8\pi}{\Omega}\right)^{\frac{1}{2}}\sum_s\alpha_s\mathrm{Q}_s\sin\gamma_s, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

5. *Phase Space for Atom and Radiation.*

It is easy to form the phase space for an assembly when the Hamiltonian of the system is known. We, therefore, in the following find out the latter for a system consisting of an electron belonging to an atom, and radiations of different frequencies associated with it.

Now the force acting on a moving electron may be written in the form

$$\mathrm{F}=-e\,\mathrm{grad}\,\mathrm{V}-\frac{e}{c}\,.\,\frac{d\mathrm{A}}{dt}+\frac{e}{c}(v\;\mathrm{curl}\;\mathrm{A}), \quad . \quad . \quad . \quad . \quad (28$$

where V and A are scalar and the vector potential and v the velocity. Therefore the equations of motion in the Lagrangian form, become

$$\left.\begin{array}{l}\frac{d}{dt}\cdot\frac{\partial L}{d\dot{q}_1}-\frac{\partial L}{dq_1}=0,\\ \text{Similarly for } q_2,\ q_3,\end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

where the Lagrangian function

$$L=\tfrac{1}{2}mv^2-eV+\frac{e}{c}(v.A). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29.1)$$

Again the total energy of the electron is.

$$W_2=\sum_{s}^{1,3} p_s\dot{q}_s-L, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where

$$p_s=\frac{\partial L}{\partial \dot{q}_s}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30.1)$$

On substituting for L and p_s in (30) from (29.1) and (30.1) respectively, we get

$$W_2=\frac{1}{2m}\sum_{s}^{1,3}\left(p_s-\frac{e}{c}A_s\right)^2+eV. \quad . \quad . \quad . \quad . \quad . \quad (30.2)$$

Thus the total energy of the system consisting of electrons and associated radiations (*vide* (26.4) and (30.2)) is

$$W=\frac{1}{2m}\left(p-\frac{e}{c}A\right)^2+eV+\sum_s(\tfrac{1}{2}P_s^2+2\pi\nu_s^2Q_s^2), \quad . \quad . \quad . \quad (31)$$

where p and A stand for the respective resultant. From the above one can easily form the phase space for the mutually interacting combined system. But it is not equal to the sum of the phase spaces for the non-interacting components. This is obvious from (31.1). Now it readily follows from (31) that

$$W=W_0-\frac{e}{mc}(p.A)+\frac{e^2}{2mc^2}.A^2, \quad . \quad . \quad . \quad . \quad . \quad (31.1)$$

where W_0 is the total energy of electrons and radiations when considered as not mutually interacting. The above relation, it may be noted, is also valid for the relativisitic case. However, it is not necessary to prove it, since the formula is well known after the work of Dirac. In the present treatment its important bearing on the phase space is only emphasized.

It is interesting to note that the second term on the right-hand side of (31.1) involves the coordinates of the electron as well as radiation. So it may be called the *coupling potential* between radiation and matter.

The third term, however, is much smaller than the second, and involves only the coordinates of radiation. Now in terms of the coordinates introduced in (26·3) they may be written as

$$\mathrm{U}_1=-\frac{e}{m}\left(\frac{8\pi}{\Omega}\right)^{\frac{1}{2}}\sum_s(p\,.\,\alpha_s)\mathrm{Q}_s\sin\gamma_s \quad . \quad . \quad . \quad . \quad (32)$$

and

$$\mathrm{U}_2=\frac{4\pi e^2}{\Omega m}\sum_{s\sigma}(\alpha_s\,.\,\alpha\ )\mathrm{Q}_s\mathrm{Q}_\sigma\sin\gamma_s\sin\gamma_\sigma \quad . \quad . \quad . \quad . \quad . \quad (33)$$

6. *Emission and Absorption of Radiation by an Atom.*

It is evident that in case of emission or absorption the interaction potential U_1 is effective, U_2 being negligible. Now, as pointed out before, we suppose that just at the time of emission or absorption a small p-volume surrounding the electron taking part in the process acquires viscosity and that the average coupling potential $(\overline{\mathrm{U}}_1)$ vanishes within this region. On the other hand, just outside the region the viscosity is zero and $\overline{\mathrm{U}}_1$ is finite. Thus we have for the χ_2-equation in q-space

$$\Delta\chi_2+\frac{8\pi^2m}{h^2}\left(\mathrm{E}-\mathrm{V}+\frac{b^2h^2}{4\pi^2\mathrm{E}}\right)\chi_2=0,\ \text{inside } (\textit{vide}\ \text{Eq. } (23.1)),$$

$$\Delta\chi_2+\frac{8\pi^2m}{h^2}(\mathrm{E}-\mathrm{V}+\overline{\mathrm{U}}_1)\chi_2=0,\ \text{outside.}$$

Since at the boundary χ_2 (inside) $=\chi_2$(outside), we have from the above

$$b^2=\frac{4\pi^2}{h^2}\,.\mathrm{E}\overline{\mathrm{U}}_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

Now on transforming (32) with the help of (15), we find that for transition from the mth to the nth eigenstate

$$\overline{\mathrm{U}}_1=eb\left(\frac{8\pi}{\Omega}\right)^{\frac{1}{2}}\sum_s\frac{mn}{(q\,.\,\alpha_s)\mathrm{Q}_s\sin\gamma_s} \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

On combining (34) and (35) we get

$$b=\frac{4\pi^2e}{h^2}\,.\mathrm{E}\left(\frac{8\pi}{\Omega}\right)^{\frac{1}{2}}\sum_s\frac{mn}{(q\,.\,\alpha_s)\mathrm{Q}_s\sin\gamma_s}. \quad . \quad . \quad . \quad . \quad . \quad (34.1)$$

Now, since

$$\mathrm{E}_m-\mathrm{E}_n=h\nu_{mn},$$

$$b_m-b_n\ =b_{mn},$$

we have from (34.1)

$$b_{mn}=\frac{4\pi^2e}{h}\left(\frac{8\pi}{\Omega}\right)^{\frac{1}{2}}\nu_{mn}\sum\frac{mn}{(q\,.\,\alpha_s)\mathrm{Q}_s\sin\gamma_s}, \quad . \quad . \quad . \quad . \quad (34.2)$$

where q and Q_s denote the coordinates of electrons and radiation $h\nu_s$. It is, of course, assumed that the energy released by $m\rightarrow n$ transition is $h\nu_{mn}=\Sigma h\nu_s$.

Now the average values of q and Q_s may be obtained as follows :—At any time t we have, since only the χ_2-wave is damped,

$$q=\int q\chi_2^m e^{-b_m t}e^{\frac{2\pi i\mathrm{E}_m}{h}.t}.\chi_1^n e^{-\frac{4\pi i\mathrm{E}_n(\mathrm{kin.})}{h}t}.d\tau$$

$$=e^{-b_m t}e^{2\pi i\nu_{mn}t}\int q\chi_2^m\chi_1^n\,d\tau.e^{2\pi i\nu_n t}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

where $\nu_n=\mathrm{E}_n-2\mathrm{E}_n(\mathrm{kin},)$. On dropping the time-factor $e^{2\pi i\nu_n t}$ in the integrand, for reasons explained before, we get from the above

$$\bar{q}=e^{-b_m t}e^{2\pi i\nu_{mn}t}.q_{mn}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36.1)$$

where q_{mn} denotes the integral without the time factor. Similarly it may be shown that for oscillators

$$\overline{\mathrm{Q}_s}=e^{-2\pi i\nu_s t}\int \mathrm{Q}_s\chi_2^{m_s}\chi_1^{n_s}\,d\tau$$

$$=e^{-2\pi i\nu_s t}\left(\frac{h}{8\pi^2\nu_s}\right)^{\frac{1}{2}}.\left.\begin{matrix}(n_s+1)^{\frac{1}{2}} & \text{if } m_s=n_s+1,\\ n_s^{\frac{1}{2}} & \text{if } m_s=n_s-1,\end{matrix}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

where m_s and n_s are the numbers of s-type quanta (or oscillators) before and after the transition. On substituting the above values of $\bar{g}$ and Q_s in (34.2) we get

$$b_{mn}=\frac{4\pi^2 e}{(\pi h\Omega)^{\frac{1}{2}}}.\nu_{mn}\sum_s\frac{1}{\nu_s^{\frac{1}{2}}}(q_{mn}.\alpha_s)\sin\gamma_s.\frac{(n_s+1)^{\frac{1}{2}}}{n_s^{\frac{1}{2}}}.e^{-b_m t}.e^{2\pi i(\nu_{mn}-\nu_s)t}. \quad (38)$$

Or, since $\sum\limits_m b_{mn}=b_n$, we have

$$b_n=\frac{4\pi^2 e}{(\pi h\Omega)^{\frac{1}{2}}}\sum_{m,s}\frac{\nu_{mn}}{\nu_s^{\frac{1}{2}}}(q_{mn}.\alpha_s)\sin\gamma_s.\frac{(n_s+1)^{\frac{1}{2}}}{n_s^{\frac{1}{2}}}.e^{-b_m t}e^{2\pi i(\nu_{mn}-\nu_s)t}\;. \quad (38.1)$$

In the same way we have

$$b_m=\frac{4\pi^2 e}{(\pi h\Omega)^{\frac{1}{2}}}\sum_{n,s}\frac{\nu_{mn}}{\nu_s^{\frac{1}{2}}}(q_{mn}.\alpha_s)\sin\gamma_s\frac{(n_s+1)^{\frac{1}{2}}}{n_s^{\frac{1}{2}}}.a_n e^{-2\pi i(\nu_{mn}-\nu_s)t}. \quad . \quad (38.2)$$

Let us consider an atom in the mth excited state at $t=0$. Suppose for simplicity that initially there is no radiation in the phase space. Thus in case of emission of $h\nu_s$, $m_s=n_s-1$ with $n_s=1$, while in case of the reverse process of absorption $m_s=n_s+1$, $n_s=0$. However, in any case we have from (38.1) and (38.2), since only two states are considered,

$$b_n=\frac{4\pi^2 e}{(\pi h\Omega)^{\frac{1}{2}}}\nu_{mn}\sum_s\frac{1}{\nu_s^{\frac{1}{2}}}(q_{mn}.\alpha_s)\sin\gamma_s\,e^{-b_m t}\,e^{2\pi i(\nu_{mn}-\nu_s)t}, \quad . \quad (38.3)$$

$$b_m=\frac{4\pi^2 e}{(\pi h\Omega)^{\frac{1}{2}}}\nu_{mn}\sum_s\frac{1}{\nu_s^{\frac{1}{2}}}(q_{mn}.\alpha_s)\sin\gamma_s a_n\,e^{-2\pi i(\nu_{mn}-\nu_s)t}. \quad . \quad . \quad (38.4)$$

Since b_n is not very great and t is small, we may take approximately $b_n = \dot{a}_n$. Hence, on integrating and remembering, that $a_n = 0$ at $t = 0$, we have from (38.3)

$$a_n = \frac{4\pi^2 e}{(\pi h \Omega)^{\frac{1}{2}}} \cdot \nu_{mn} \sum_s \frac{1}{\nu_s^{\frac{1}{2}}} (q_{mn} \cdot \alpha_s) \sin \gamma_s \frac{e^{[-b_m + 2\pi i(\nu_{mn} - \nu_s)]t} - 1}{-b_m + 2\pi i(\nu_{mn} - \nu_s)} \quad . \; . \quad (38.5)$$

On substituting the above value of a_n in (38.4), we have

$$b_m = \frac{16\pi^3 e^2}{h\Omega} \cdot \nu_m^2 \, \sum_s \frac{1}{\nu_s} (q_{mn} \cdot \alpha_s)^2 \sin^2 \gamma_s e^{-b_m t} \cdot \frac{e^{[b_m - 2\pi i(\nu_{mn} - \nu_s)]t} - 1}{b_m - 2\pi i(\nu_{mn} - \nu_s)} \; . \quad (39)$$

The sum may evidently be replaced by an integral over ν_s after multiplying by the factor $\frac{8\pi\Omega}{c^3} \nu_s{}^2 \, d\nu_s$, which gives the number of radiations in the phase space between frequencies ν_s and $\nu_s + d\nu_s$. Also we may take the average values over all phases, directions, and polarizations, viz.,

$$\overline{(q_{mn} \cdot \alpha_s)^2} = \tfrac{1}{3} q^2. \quad \overline{\sin^2 \gamma_s} = \tfrac{1}{2}. \quad . \; . \; . \; . \; . \; . \quad (39.1)$$

Hence we get from (39), since $e^{-b_m t} \to 1$,

$$b_m = \frac{64\pi^4 e^2}{3hc^3} \nu_{mn}^2 q^2 \int_0^\infty \frac{e^{[b_m - 2\pi i(\nu_{mn} - \nu_s)]t} - 1}{b_m - 2\pi i(\nu_{mn} - \nu_s)} \nu_s \, d\nu_s. \quad . \; . \quad (39.2)$$

Since b_m may be neglected compared with ν's, the above integral has the value $\frac{\nu_{mn}}{2}$. It is evident that the value of the integral is concentrated near $\nu_s \sim \nu_{mn}$. Thus it follows that although we have started from the assumption that radiations of all frequencies ν_s's may be emitted, the probability of emission is negligible except for the value $\nu_s \sim \nu_{mn}$. We have now from (39.2)

$$b_m = \frac{32\pi^4 e^2}{3hc^3} \cdot \nu_{mn}^3 q^2. \quad . \; . \; . \; . \; . \; . \; . \; . \quad (40)$$

The reciprocal of b_m gives the mean life (τ) of the stationary state. Hence from (40) we have

$$\tau = \frac{3hc^3}{32\pi^4 e^2 \nu^3 q^2} \; . \quad . \; . \; . \; . \; . \; . \; . \; . \quad (41)$$

It may be noted that Einstein's quantum statistical value is double the above. Dirac, however, obtains for the mean life half of the above value, apparently because he takes both the q- and p-components of the phase fluid to be viscous. But, as has been pointed out before, only the p-component should be taken viscous in case of emission or absorption of radiation.

7. *Propagation of Radiation through Empty Space.*

Let us consider two atoms A and B in an assembly at a distance r apart. Suppose that at time $t=0$ A is in an excited state (m) and B in the normal state (n'). After a certain time A goes over to the normal nth state, emitting radiations which are absorbed by B going to the m'-excited state. Now we have from (39), on integrating with respect to time, and remembering that the constant of integration vanishes because of the condition $a_m=0$ for $t=\infty$,

$$a_m=\frac{16\pi^3e^2}{h\Omega}\cdot\nu_{mn}^2\,\mathop{\Sigma}_s\frac{1}{\nu_s}(q_{mn}\cdot\alpha_s)^2\sin^2\gamma_s\cdot\frac{1}{b_m-2\pi i(\nu_{mn}-\nu_s)}$$
$$\times\left\{-\frac{e^{-2\pi i(\nu_{mn}-\nu_s)t}}{2\pi i(\nu_{mn}-\nu_s)}-\frac{e^{-b_mt}}{b_m}\right\}.\quad . \quad . \quad (42)$$

Because b_m is great the second term within brackets in (42), decreasing exponentially with time, becomes negligible in a short time. Therefore

$$a_m=\frac{16\pi^3e^2}{h\Omega}\cdot\nu_{mn}^2\,\mathop{\Sigma}_s\frac{1}{\nu_s}(q_{mn}\cdot\alpha_s)^2\sin^2\gamma_s\cdot\frac{1}{b_m-2\pi i(\nu_{mn}-\nu_s)}\cdot\frac{e^{2\pi i(\nu_s-\nu_{mn}}}{2\pi i(\nu_s-\nu_{mn})}\cdot$$
$$. \quad . \quad . \quad (42.1)$$

Since the q-space is not viscous, a_m is unity in this part of the phase space. So (42.1) gives the probability of emission of radiation by the atom A. Similarly we have for the probability of absorption of the radiation by the atom B

$$a_{n'}=\frac{16\pi^3e^2}{h\Omega}\nu_{m'n'}^2\,\mathop{\Sigma}_{s'}\frac{1}{\nu_{s'}}(q_{m'n'}\cdot\alpha_{s'})^2\sin^2\gamma'_{s'}\cdot\frac{1}{b_{n'}-2\pi i(\nu_{m'n'}-\nu_{s'})}\cdot\frac{e^{2\pi i(\nu_{s'}-\nu_{m'n'})t}}{2\pi i(\nu_{s'}-\nu_{m'n'})}\cdot$$
$$. \quad . \quad . \quad (43)$$

Hence the joint probability of emission by A and absorption by B is given by $a_m\cdot a_{n'}=a_{mn'}$, where

$$a_{mn'}=\frac{2^8\pi^6e^4}{h^2\Omega^2}\nu_{mn}^2\nu_{m'n'}^2\,\mathop{\Sigma}_s\mathop{\Sigma}_{s'}\frac{1}{\nu_s\nu_{s'}}(q_{mn}\cdot\alpha_s)^2\cdot(q_{m'n'}\cdot\alpha_{s'})^2$$
$$\times(\sin\gamma_s\sin\gamma'_{s'})^2\cdot\frac{1}{b_m-2\pi i(\nu_{mn}-\nu_s)}\cdot\frac{1}{b_{n'}-2\pi i(\nu_{m'n'}-\nu_{s'})}$$
$$\times\frac{e^{2\pi i(\nu_s{}'-\nu_{mn})t}}{2\pi i(\nu_s-\nu_{mn})}\cdot\frac{e^{2\pi i\,\nu_{s'}-\nu_{m'n'})t}}{2\pi i(\nu_{s'}-\nu_{m'n'})}.\quad . \quad . \quad (44)$$

For simplicity we suppose that the atoms A and B are on the x-axis at a distance r, and that the vectors q_{mn}, $q_{m'n'}$ are reduced to the y-components. We have then for the average value

$$\overline{(q_{mn}\cdot\alpha_s)^2\cdot(q_{m'n'}\cdot\alpha_{s'})^2}=\tfrac{1}{4}q^2q'^2).\quad . \quad . \quad . \quad . \quad . \quad (44.1)$$

Also we have

$$\left.\begin{aligned}\gamma_{s'}&=\gamma_{s'}+\frac{2\pi\nu_{s'}}{c}r\cos\theta_{s'},\\ \gamma_s&=\gamma'_s+\frac{2\pi\nu_s}{c}r\cos\theta_s,\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44.2)$$

θ_s, $\theta_{s'}$ being the angles between the x-axis and the directions of s- and s'-components of radiation. Thus the average value over all phases and directions,

$$\overline{\sin\gamma_s\sin\gamma_{s'}}=\frac{c}{4\pi\nu_{s'}r}\sin\frac{2\pi\nu_{s'}r}{c}=\frac{c}{4\pi\nu_s r}\sin\frac{2\pi\nu_s r}{c}. \quad . \quad . \quad (44.3)$$

On using the average values given in (44.1) and (44.3) we get from (44)

$$a_{mn'}=\frac{4\pi^4e^4c^2}{h^2\,\Omega^2r^2}q^2q'^2\nu^2_{mn}\nu_{m'n'}\sum_s\sum_{s'}\frac{1}{\nu^2_s\nu^2_{s'}}\frac{1}{b_m-2\pi i(\nu_{mn}-\nu_s)}$$

$$\times\frac{1}{b_{n'}-2\pi i(\nu_{m'n'}-\nu_{s'})}\sin\frac{2\pi\nu_s r}{c}\sin\frac{2\pi\nu_{s'}r}{c}\cdot\frac{e^{2\pi i(\nu_s-\nu_{mn})t}}{2\pi i(\nu_s-\nu_{mn})}$$

$$\times\frac{e^{2\pi i(\nu_{s'}-\nu_{m'n'})t}}{2\pi i(\nu_{s'}-\nu_{m'a'})}. \quad . \quad . \quad (44.4)$$

As before, the summations may be replaced by integrations. Again, due to the factors $(\nu_s-\nu_{mn})$ and $(\nu_{s'}-\nu_{m'n'})$ in the denominator, the values of the integrands are concentrated in the neighbourhoods of ν_{mn} and $\nu_{m'n'}$ for the variables ν_s and $\nu_{s'}$ respectively. We may therefore take the factors $\frac{1}{b_m-2\pi i(\nu_{mn}-\nu_s)}$ and $\frac{1}{b_{n'}-2\pi i(\nu_{m'n'}-\nu_{s'})}$, which change regularly, out of the integrals after putting $\nu_s=\nu_{mn}$ and $\nu_{s'}=\nu_{m'n'}$. Therefore we have from (44.4)

$$a_{mn'}=\frac{2^8\pi^6e^4}{b_mb_{n'}h^2c^4r^2}q^2q'^2\nu^2_{mn}\nu^2_{m'n'}\int_0^\infty\int_0^\infty\frac{e^{2\pi i(\nu_s-\nu_{mn})t}}{2\pi i(\nu_s-\nu_{mn})}\sin\frac{2\pi\nu_s r}{c}$$

$$\times\frac{e^{2\pi i(\nu_{s'}-\nu_{m'n'})t}}{2\pi i(\nu_{s'}-\nu_{m'n'})}\sin\frac{2\pi\nu_{s'}r}{c}d\nu_s\,d\nu_{s'}. \quad . \quad . \quad (44.5)$$

Because ν's are great, the integrals in (44.5) may be reduced to the form

$$\int_{-\infty}^{+\infty}\frac{e^{2\pi i\xi t}}{2\pi i\xi}\sin\frac{2\pi r}{c}(\xi+\nu)\,d\xi,$$

which becomes, after the imaginary term is neglected,

$$\frac{1}{2\pi}\int_{-\infty}^{+\infty}\frac{1}{\xi}\sin 2\pi\xi t\sin\frac{2\pi r}{c}(\xi+\nu)\,d\xi=0 \quad \text{if } t<\frac{r}{c}$$

$$=\tfrac{1}{2}\sin\frac{2\pi\nu r}{c} \quad \text{if } t>\frac{r}{c}.$$

Thus we have from (44.5), since $t > \frac{r}{c}$,

$$a_{mn'} = \frac{1}{r^2}\left[\frac{8\pi^3 e^2}{c^2 h}\nu_{mn}q\ \nu_{m'n'}q'\right]^2 \cdot \frac{1}{b_m b_{n'}} \sin\frac{2\pi\nu_{mn}r}{c} \sin\frac{2\pi\nu_{m'n'}r}{c}, \quad . \quad (45)$$

showing that the probability is inversely proportional to the square of the distance r. Thus the theory explains propagation of radiation and the law of decrease of intensity with distance.

8. *Scattering of Radiation by Free Electron.*

We suppose that the volume immediately surrounding an electron temporarily acquires negative viscosity relative to the region outside; and so the radiation ($h\nu_s$) enters the given volume. As a result the phase density rapidly increases and radiations $\sum\limits_{\sigma} h\nu_\sigma$ are scattered in all directions. The scattering is evidently due to positive viscosity developed in the given volume. It is obvious that for the above mechanism of scattering we should take both the χ_1- and χ_2-waves to be damped. We further suppose that, due to scattering, the momentum of the electron is not changed on an average. Therefore U_1 in (32) cannot be taken as the interaction potential in the present case. It should be U_2, given in (33). Now, assuming as before that within the given volume surrounding the scattering electron $\overline{U}_2 = 0$ and $b \neq 0$, but outside it $\overline{U}_2 \neq 0$ and $b = 0$, we have, corresponding to (34),

$$b^s_1 = b^s_2 = \frac{2\pi}{h}\sqrt{h\nu_s \overline{\overline{U^s_2}}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

where b^s_1 and b^s_2 are respectively the damping coefficients in the q- and p-spaces for s-type radiation. Now from (33) we have, on putting, $s = \sigma$

$$\overline{U^s_2} = \frac{8\pi e^2}{\Omega m} \cdot \overline{\alpha^2_s Q^2_s \sin^2\gamma_s}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

The primary radiation is of frequency ν_s. Let its number density be $\frac{n_s}{\Omega}$. Due to scattering it is evidently changed to $\frac{n_s - 1}{\Omega}$. Hence we have from (37)

$$Q_s = e^{-2\pi i\nu_s t}\left(\frac{hn_s}{8\pi^2\nu_s}\right)^{\frac{1}{2}}.$$

As $\overline{Q_s^2} = (Q_s)^2$ approximately, we have from (47)

$$\overline{U^s_2} = \frac{he^2}{\pi\Omega m} \cdot \frac{n_s}{\nu_s} . \overline{\alpha^2_s \sin^2\gamma_s} . e^{-4\pi i\nu_s t}. \quad . \quad . \quad . \quad . \quad . \quad (47.1)$$

Similarly we have for the scattered radiation of frequency ν_σ, since in this case the number density changes from $0 \rightarrow 1/\Omega$,

$$\overline{U^\sigma_2} = \frac{he^2}{\pi\Omega m} \cdot \frac{1}{\nu_\sigma} . \overline{\alpha^2_\sigma \sin^2\gamma_\sigma}\, e^{4\pi i\nu_\sigma t}. \quad . \quad . \quad . \quad . \quad (47.2)$$

Now from (46) and (47.1) we get

$$b_1^s = b_2^s = 2\pi \left\{ \frac{n_s e^2}{\pi \Omega m} \cdot \overline{\alpha_s^2 \sin^2 \gamma_s} \right\}^{\frac{1}{2}} e^{-2\pi i \nu_s t}. \quad . \quad . \quad . \quad (48)$$

In the same way we have

$$b_1^\sigma = b_2^\sigma = 2\pi \left\{ \frac{e^2}{\pi \Omega m} \cdot \overline{\alpha_\sigma^2 \sin^2 \gamma_\sigma} \right\}^{\frac{1}{2}} e^{2\pi i \nu_\sigma t}. \quad . \quad . \quad . \quad (48.1)$$

Since $b_1^s = \dot{a}_1^s$, we obtain from (48) on integrating with respect to t,

$$a_1^s = 2\pi \left\{ \frac{n_s e^2}{\pi \Omega m} \cdot \overline{\alpha_s^2 \sin^2 \gamma_s} \right\}^{\frac{1}{2}} \frac{e^{-2\pi i \nu_s t}}{-2\pi i \nu_s}. \quad . \quad . \quad . \quad . \quad (49)$$

Hence the rate of scattering for the q-space is

$$b_1^{s\sigma} = a_1^s b_1^\sigma = \frac{2e^2}{\Omega m} \left\{ \overline{\alpha_s^2 \alpha_\sigma^2 \sin^2 \gamma_s \sin^2 \gamma_\sigma} \right\}^{\frac{1}{2}} n_s^{\frac{1}{2}} \cdot \frac{e^{-2\pi i(\nu_s - \nu_\sigma)t}}{-i\nu_s}. \quad . \quad . \quad (50)$$

Now as $b_1^{s\sigma} = \dot{a}_1^{s\sigma}$ we have on integrating (50) with respect to time and $a_1^{s\sigma}$ is zero at $t=0$,

$$a_1^{s\sigma} = \frac{e^2}{\pi \Omega m} \left\{ \overline{\alpha_s^2 \alpha_\sigma^2 \sin^2 \gamma_s \sin^2 \gamma_\sigma} \right\}^{\frac{1}{2}} \cdot n_s^{\frac{1}{2}} \cdot \frac{1 - e^{-2\pi i(\nu_s - \nu_\sigma)t}}{\nu_s(\nu_s - \nu_\sigma)} \quad . \quad . \quad (51)$$

Similarly

$$a_2^{s\sigma} = \frac{e^2}{\pi \Omega m} \left\{ \overline{\alpha_s^2 \alpha_\sigma^2 \sin^2 \gamma_s \sin^2 \gamma_\sigma} \right\}^{\frac{1}{2}} \cdot n_s^{\frac{1}{2}} \cdot \frac{1 - e^{2\pi i(\nu_s - \nu_\sigma)t}}{\nu_s(\nu_s - \nu_\sigma)} \quad . \quad . \quad (51.1)$$

It is evident that the total number of quanta scattered in any component is

$$N^s = \sum_\sigma a_1^{s\sigma} \cdot a_2^{s\sigma}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51.2)$$

As in the preceding sections, we replace the summation by integration and substitute for $(\alpha_s \alpha_\sigma)^2$, $\sin^2 \gamma_s$, $\sin^2 \gamma_\sigma$ their mean values, $\frac{1}{3}$, $\frac{1}{2}$, and $\frac{1}{2}$. We have

$$N^s = \frac{8e^4}{3\pi \Omega m^2 c^3} \cdot \frac{n_s}{\nu_s^2} \int_0^\infty \frac{\sin^2 \pi(\nu_s - \nu_\sigma)t}{(\nu_s - \nu_\sigma)^2} \nu_\sigma^2 \, d\nu_\sigma. \quad . \quad . \quad . \quad . \quad (52)$$

Because the value of the integrand is concentrated near $\nu_\sigma = \nu_s$ we may take ν_σ^2 out of the integral as ν_s^2. The remaining integral can be easily transformed to the standard form,

$$\int_{-\infty}^{+\infty} \frac{\sin^2 kx}{x^2} \, dx = \pi k.$$

Therefore we have from (52)

$$N^s = \frac{8\pi}{3} \cdot \frac{e^4}{c^3 m^2} \cdot \frac{n_s}{\Omega} \cdot t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

which is the well-known Thomson formula for the scattering of radiation, derived from the classical theory.

Physical Laboratory,
Presidency College, Calcutta, India.
January 1939.

XXI. *The Specific Heat of Water and of Heavy Water.*

By A. H. Cockett and Allan Ferguson *.

[Received December 27, 1939.]

Introduction.

Through the kindness of Imperial Chemical Industries, Ltd., we have been supplied with a specimen of heavy water of 99·2 per cent. purity.

This communication describes determinations of the specific heat of ordinary and of heavy water *at* different temperatures over a range of approximately 10°–50° C.

The method employed, which, with the present apparatus, is capable of a consistency of at least 0·1 per cent., is specially suited for experiments on liquids obtainable only in relatively small quantities, inasmuch as a determination can be made on a sample of liquid of volume 300–400 c.c.

Method.

Although the methods that have already been used to measure the temperature variation of the specific heat of water are many and varied, there is one feature common to them all. In every important case it is the mean specific heat over a given temperature range that is measured, and from this mean value the actual specific heat at the intermediate temperature is calculated. The temperature range over which the mean is measured may be as large as from 11° to 18° C., as was the fundamental interval in the experiments of Ludin, or it may be as small as one or two degrees Centigrade, as it was in the experiments of Callendar and Barnes.

The method adopted in the present investigation is one by which the specific heat of a liquid *at* any desired temperature may be measured. The method is unorthodox in that the heat losses, instead of being diminished as far as possible, so that they necessitate the making of merely a small correction to the observations, are measured as one of the fundamental parts of the experimental procedure.

The method was used by Ferguson and Miller (1) in an investigation of the temperature variation of the specific heats of organic liquids, and in principle it is as follows. A known quantity of the liquid is heated electrically, the power supplied being measured and so adjusted that the difference between the temperature of the liquid and that of the surrounding atmosphere is maintained constant. That is, for a given value of this temperature difference θ, the rate of loss of heat to the surroundings is measured, and it remains to find how this depends on θ, the heat capacity of the calorimeter and its accessories and that of the liquid. This is achieved by plotting a cooling curve, the power of the method lying in the simple relation which has been found to exist between the rate of cooling, $-\frac{d\theta}{dt}$, and θ and s, the specific heat of the liquid. As the result of a large number of experiments, made at very different temperatures, it is found that this may be expressed in the equation

---

* Communicated by the Authors.

$$-\frac{d\theta}{dt}=K\theta^n .$$

The rate of loss of heat is therefore

$$(Ms+W)K\theta^n,$$

and this must be equated to the power supplied, which is measured in terms of the resistance R of the heating coil, and the electromotive force E across it. The equation which expresses the relation between the observed quantities is, therefore,

$$(Ms+W)K\theta^n=\frac{E^2}{RJ},$$

where the symbols have their usual meaning.

Of the quantities involved in this equation, M, E, R, and θ are the directly observed ones, while K and n are found by a graphical method, or its equivalent, and W is found by an experiment at the temperature at which s is defined as equal to unity. s at any other temperature is then calculable.

Experimental.

The essential parts of the apparatus used in these experiments are :

1. Calorimeter.
2. Heating Coil.
3. Thermostat.
4. Thermometer.
5. Source of current.
6. Potentiometer.
7. Time-piece.

These are separately described below.

The Calorimeter.

The calorimeter can best be described by reference to fig. 1. It consisted of a cylindrical copper vessel A, about twelve centimetres in height, eight in diameter, and one and a half millimetres thick. It was provided with a machined flange B to which the lid C was attached by means of six screws. The lid, also of copper, carried the shaft D with the stirrer blades, all of stainless steel, the thermometer F and the heating coil with its potential leads G. In order to reduce to a minimum the heat produced by the stirrer in its bearings H, a ball bearing was used. The calorimeter stood on three short legs K, of each of which the upper half consisted of copper, while the lower half was of bakelite. A short length of bakelite rod was inserted in the shaft of the stirrer rod.

For the purpose of a previous experiment the calorimeter had been thinly chromium plated, and this plating was not removed.

The Heating Coil.

The material chosen for the heating coil was constantan, on account of its negligible temperature coefficient of resistance. Two thick copper rods acted as current-carrier leads, and were soldered to two of four terminal screws carried by, but insulated from, the lid of the calorimeter. Two thinner copper rods, connected to the junctions of the ends of the coil with the current leads, served as potential leads and were soldered to the remaining two screws in the lid. The constantan coil and the copper leads were insulated by means of a baked-on layer of bakelite cement. This was made as thin as possible, consistent with efficient insulation, in order to prevent lag in the rate of transfer of heat from the wire to the liquid, bakelite being a poor conductor of heat.

The resistance of the coil was about 0·1 ohm.

The Thermostat.

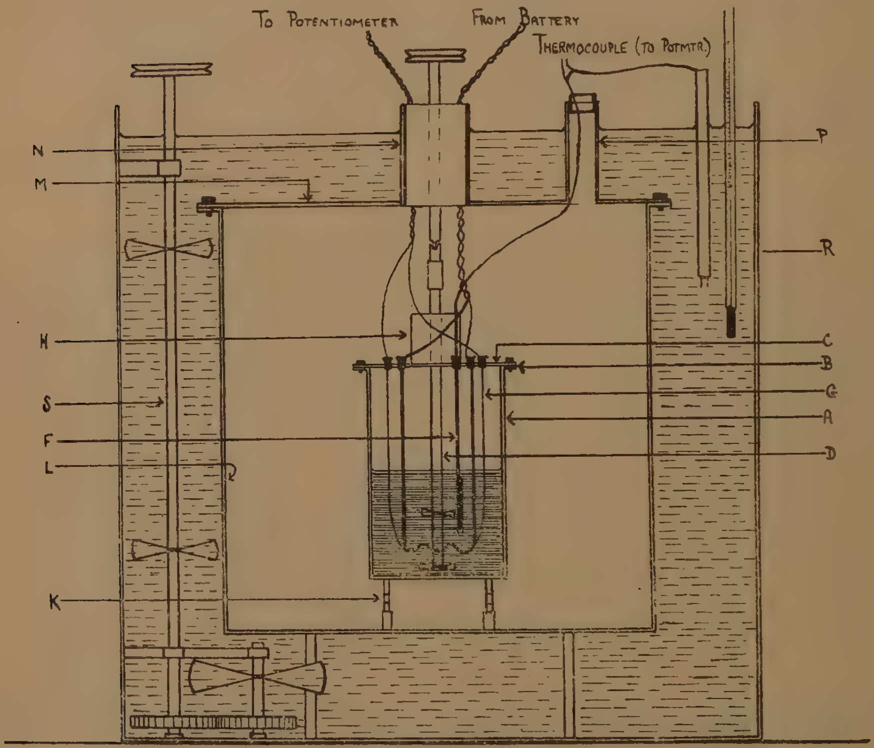
In order to define accurately the temperature of the surroundings, and in view of the fact that it was anticipated that this would be in the neighbourhood of 0° C., it was decided to arrange for the calorimeter to be enclosed in a vessel completely immersed in water. For this purpose the thermostat was of the construction shown in fig. 1.

The calorimeter is indicated at A, enclosed in the vessel L, which was thirty centimetres in height and diameter. In it the three legs of the calorimeter fitted into corresponding slots, so that the position occupied by the calorimeter was always exactly the same. The vessel L was closed by a copper disk M which was attached by means of eight screws to a flange at the top of L. An oiled leatheroid gasket proved very successful in ensuring a water-tight seal.

The stirrer of the calorimeter was actuated by means of a shaft passing down the tube N in the centre of the disk, operating the shaft D through a dog clutch. A second tube P served for the introduction of leads to the thermocouple.

The vessel L rested on five pegs soldered outside its base which fitted in corresponding tubes in the large tank R, the space between the two, seven and a half centimetres wide, being filled with water or, more often, a mixture of ice and water. As it was considered essential to obtain thorough stirring of this ice-water mixture in order to ensure equality of temperature throughout, stirring-paddles were fixed on the shaft S, a fourth and larger blade being placed nearer the centre of the tank, in the space below L. This one was operated from S through gearing.

Fig. 1.



Calorimeter and thermostat.

A simple calculation gives the volume of water required to fill the space between the walls L and R of the thermostat as 45 litres. Even when the temperature was very nearly 0° C., this water was so thoroughly stirred that experiments made on several occasions failed to reveal any difference greater than 0·01° C., the minimum observable, in the temperature of different parts of the tank, whether measured by means of a mercury in glass thermometer or a thermocouple.

The exterior of the tank was lagged with a single layer of corrugated packing cardboard.

Thermometers.

In the measurement of temperature, two quantities were required with accuracy ; first, the temperature of the contents of the calorimeter, and second, the difference between this temperature and that of the thermostat.

The temperature of the thermostat was measured by means of a mercury-in-glass thermometer obtained from B. Black & Son, graduated in tenths of a degree Centigrade over the range 0°–100° C. This had been certified by the National Physical Laboratory, and was of a high degree of accuracy. In the majority of the experiments the temperature of the thermostat was, within a few tenths of a degree, 0° C., and it was principally this region of the thermometer scale that was used, although on a few occasions the thermostat was used at temperatures in the region of 10–15° C.

In view of the fact that. as has been described above, the calorimeter was totally enclosed. it would have been impossible to use a mercury-in-glass thermometer to measure its temperature, and it was therefore decided to use a thermocouple.

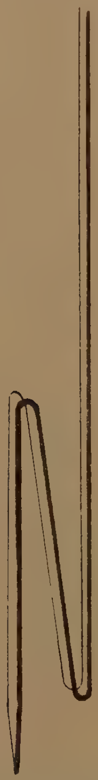
The metals chosen to constitute the thermocouple were copper and constantan. The importance of ensuring equality of temperature between the junction and the liquid necessitated a low rate of conduction of heat along the wires constituting the couple, which was achieved by the use of thin wires (gauge 30 S.W.G.). As a further safeguard, the length of wire immersed in the liquid was made as great as possible without unduly increasing the water equivalent. To confine the couple to a reasonably small space the wires were bent into an " N " shape, as indicated in fig. 2, the total length of wire immersed being about 8–10 centimetres.

The smallest potential difference that the potentiometer was capable of measuring was one microvolt, although fractions down to about one-quarter could be estimated with reasonable accuracy. To measure accurately to 0·01° C. it was therefore considered necessary to have a thermo-electric-power of at least 100 microvolts per degree Centigrade. Since the e.m.f. of a copper-constantan thermocouple is of the order 40 microvolts per degree Centigrade, this thermo-electric power was attained by using three thermo-junctions in series. giving about 120 microvolts per degree Centigrade.

The thermocouple was sent to the National Physical Laboratory for calibration over the range 15–50° C., the conditions of immersion closely resembling those under which it was used.

In use the hot junction was immersed in the calorimeter while the cold junction was placed in the mixture of ice and water filling the thermostat. This was usually at a temperature merely a few hundredths of a degree from 0° C., and in calculating the temperature of the calorimeter the observed e.m.f. was increased by an amount equivalent to the e.m.f. of the thermocouple at this temperature.

Fig. 2.



Disposition of thermocouple.

On a few occasions the thermostat was maintained at a temperature in the neighbourhood of 10° C., and it was verified that, under these conditions, the temperature of the calorimeter could be calculated accurately from the sum of the e.m.f. at the temperature of the bath and the observed e.m.f.

The Source of Current.

The heating current was supplied initially by a battery of accumulators, and was regulated by means of resistances in series with the heating coil. Coarse adjustment was made by means of low resistance, and fine adjustments by the use of a high resistance in parallel with the low ones. The maximum fluctuations permitted in the electromotive force across the heating coil during any experiment were approximately ±30 microvolts, so that the mean was known to an accuracy greater than this, say ±20 microvolts. At the smallest currents used, about 2 amps., corresponding to 0·2 volt across the coil, this corresponds to an accuracy of one part in ten thousand.

The Potentiometer.

This was an excellent instrument made by Pye & Co., with ranges 0–2·2 volts, 0–0·22 volt, and 0–0·022 volt. On the first range the units and tenths were obtained from the dial, a contact switch bringing resistance coils into circuit, while the remaining figures were obtained from a slider moving over a uniform wire, 50 centimetres long divided into 1000 parts, numbered in tens 1 to 100, the whole length of wire being equivalent to 0·1 volt on this range. On the third range one of the smallest divisions was equivalent to one microvolt.

By means of this instrument it was possible to measure e.m.f.'s correct to five significant figures, the corrections given on the makers' calibration certificate being too small to read ; the ratio resistances were accurate to the same degree. The e.m.f.'s were compared with the e.m.f. of a Weston standard cell, which had been checked at the National Physical Laboratory.

A high resistance moving-coil galvanometer was used, whose sensitivity was such that a deflexion of 1 mm. at 2 metres was equivalent to 4 microvolts.

The Time-piece.

In plotting cooling curves the time intervals were measured by a pocket-watch, occasionally in conjunction with a stop-watch. The pocket-watch was compared frequently with Greenwich time.

Technique.

As has already been stated, the experimental procedure falls into two sections, described as " heating experiments " and " cooling curves."

A. *Heating Experiments.*

This section of the work consisted in passing a known current through the heating coil and observing the steady difference between the temperature of the calorimeter and that of the thermostat.

Initially the water was heated by passing a current of 2 amperes through an auxiliary heating coil of resistance about 3 ohms, which was used solely for the purpose of effecting rapid changes of temperature. When the desired temperature was obtained this current was switched off and the current started in the main heating coil. This was adjusted to a value deemed approximately suitable to maintain constant the existing temperature differences. It was found easy, after having obtained a little experience, to estimate this with reasonable accuracy.

TABLE I.

Observations made to determine the Power required to maintain a Constant Temperature Difference.

Period.	E.M.F. across heating coil, volts.	Thermoelectric E.M.F. (micro-volts).	
		At commencement of period.	At end of period.
11.50–11.55 a.m.	0·73543	2284	2290
11.55–12.00 noon.	0·73040	2290	2288
12.00–12.05 p.m.	0·73145	2288	2284
12.05–12.10 p.m.	0·73248	2284	2281
12.10–12.15 p.m.	0·73342	2281	2281
12.15–12.20 p.m.	0·73342	2281	2281
12.20–12.25 p.m.	0·73342	2281	2281

Temperature of thermostat	0·15° C.
Final steady temperature difference	19·67° C.
E.M.F. across heating coil	0·73342 volts.

When the current had been flowing for five minutes the temperature difference was observed and the same current was maintained constant for a further equal period, another temperature observation being then made. According to the change in the temperature difference the heating current was maintained constant, increased or decreased. After keeping this value constant for five minutes, a further temperature reading was taken and another adjustment of the current was made if necessary. By this means a value of the heating current required to maintain the temperature difference constant for fifteen minutes was obtained. Usually the final steady temperature of the calorimeter was within two or three tenths of a degree Centigrade of the required value.

Having obtained one set of observations, the current was switched on in the auxiliary coil, a new temperature was attained, and a second series of observations followed. The difference between the temperatures of the calorimeter and the thermostat was increased in this way, so that on any one day a fairly wide range of temperatures was covered, usually 10–15° C. approximately.

The readings made to determine a steady temperature difference and the current to maintain it are shown in Table I., which gives a typical set of observations.

B. *Cooling Curves.*

Having in this way built up a considerable temperature difference, the heating current was switched off and a cooling curve obtained. It was found most satisfactory for this purpose to observe the time at which the temperature of the calorimeter exceeded that of the thermostat by certain pre-chosen amounts. Accordingly the potentiometer contact was set at the appropriate e.m.f. and the instant at which the spot of light reflected from the galvanometer mirror passed through its zero position was noted. A succession of such observations made it possible to plot a cooling curve.

In order to obtain the desired accuracy of at least one part in a thousand, each cooling curve was continued over a time interval of about a half to two hours, and contained 10–15 observations, according to the rate of cooling, which depended on the initial temperature excess and the quantity of water in the calorimeter.

Calculation of Results.

The description of the experimental work has been split up into two sections corresponding to the two groups into which the experiments themselves fell. It is now necessary to link these up so that the specific heat may be calculated.

In terms of the symbols which have already been defined, we start with the two equations

$$\frac{E^2}{RJ} = -(Ms+W)\frac{d\theta}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (A)$$

corresponding to the ordinary equation of calorimetry :

"Quantity of Heat supplied=Quantity of Heat dissipated," and

$$-\frac{d\theta}{dt} = K\theta^n \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (B)$$

which expresses the march of the cooling curves.

On combining these equations we obtain

$$(Ms+W)K\theta^n=\frac{E^2}{RJ} \quad . \; . \; . \; . \; . \; . \; . \; . \quad (C)$$

in which we require to know K and n before it is possible to calculate s.

In the work of Ferguson and Miller previously referred to ⁽¹⁾ it was shown that n was equal to 5/4. Thus, a plot of $1/\theta^{n-1}$ against time, where n is given the value 5/4, gives a straight line, since on integrating the equation B, we have

$$\frac{1}{\theta^{n-1}}-\frac{1}{\theta_0^{\,n-1}}=(n-1)Kt, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (D)$$

where θ_0 is the value of θ when $t=0$.

Early in the present series of experiments it was found that this value of n did not fit the observations— a plot of $\theta^{-\frac{1}{4}}$ against t showing appreciable curvature. It was also seen that a better value of n was 1·2, but it was not known how closely this met the requirements of the observations, as it was impossible to judge sufficiently accurately from a graph whether it should have been, say 1·21 or 1·19, and a method of calculating n from the observations was sought *.

It is convenient to mention here that it is not surprising to find this difference in the values of the index appropriate to the experiments made by Ferguson and Miller, and to the present experiments. The former were made in a thermostat completely open at the top, so that convection currents set up in the air surrounding the calorimeter were able to rise vertically to what was, for all practical purposes, an infinite height. In the present work, however, the thermostat was closed at the top, and the convection currents had therefore a very restricted space in which to circulate.

The method of calculating n is as follows :—

The equation D is rewritten in the form

$$y=a-bx^m, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (E)$$

where y stands for t in equation D.

a stands for $\frac{-1}{(n-1)K}\cdot\frac{1}{\theta_0^{\,n-1}}$ and is a constant.

b stands for $\frac{-1}{(n-1)K}$, a constant.

m stands for $-(n-1)$ and is also a constant.

x stands for θ.

---

* We desire to thank Mr. J. H. Awbery for valuable suggestions.

It is clear that a knowledge of the constants b and m suffices to determine K and n.

Let the true value of the constants be a, b, and m, while a_0, b_0, and m_0 are approximate values obtained by any convenient method, say graphically, and put

$$a=a_0(1+\alpha),$$

$$b=b_0(1+\beta),$$

$$m=m_0(1+\gamma),$$

where it is assumed that α, β, γ are small.

Equation E can now be rewritten as

$$y=a_0(1+\alpha)-b_0(1+\beta)x^{m_0}.\,x^{m_0\gamma},$$

or as

$$y=a_0(1+\alpha)-b_0(1+\beta)x^{m_0}e^{m_0\gamma\log_e x}.$$

$$=a_0(1+\alpha)-b_0(1+\beta)x^{m_0}(1+m_0\gamma\log_e x)$$

provided that $m_0\gamma\log_e x$ is small.

On expanding and rearranging we have (neglecting products of the small quantities α, β, γ)

$$y-a_0+b_0x^{m_0}=a_0\alpha-b_0\beta x^{m_0}-b_0m_0\log_e x\,.\,\gamma. \quad . \quad . \quad . \quad . \quad \text{(F)}$$

In this equation all the quantities on the left-hand side are known, while on the right-hand side the unknowns are the three small quantities α, β, γ.

The observations may now be treated by the method of least squares. The terminology is further simplified by putting

$$\text{A}=a_0\alpha,\ \text{B}=-\beta,\ \text{C}=-\gamma,$$

$$\text{Y}=y-a_0+b_0x^{m_0},\ \text{X}=b_0x^{m_0}\text{ and }\text{Z}=b_0m_0\log_e x,$$

giving equation (F) as

$$\text{Y}=\text{A}+\text{BX}+\text{CZ}.$$

Applying the standard procedure we obtain three equations from which it is possible to calculate A, B, and C from the cooling curve observations. These equations are

$$\Sigma\text{Y}-p\text{A}-\text{B}\Sigma\text{X}-\text{C}\Sigma\text{Z}=0.$$

$$\Sigma\text{XY}-\text{A}\Sigma\text{X}-\text{B}\Sigma\text{X}^2-\text{C}\Sigma\text{ZX}=0.$$

$$\Sigma\text{YZ}-\text{A}\Sigma\text{Z}-\text{B}\Sigma\text{ZX}-\text{C}\Sigma\text{Z}^2=0.$$

In these p is the number of observations made for plotting the cooling curve.

Sources of Error.

In order to estimate the relative importance of various sources of error the following measurements were made, with the results stated :—

(*a*) Heating effect of stirrer. No observable change of temperature occurred over a period of one hour of continued stirring, the calorimeter being at the temperature of the thermostat.

(*b*) Efficiency of stirring. Except when changing the temperature of the liquid between the experiments, there was no observable temperature difference throughout the calorimeter.

(*c*) Loss of liquid by evaporation. When the temperature of the calorimeter and liquid was 75° C. this was 0·01 gram in three hours. At the highest temperature at which the specific heat was measured, 50° C., the rate of evaporation would be about one-third of this, *i. e.*, 0·001 gram per hour, equivalent to a loss of heat at the rate of 0·02 calorie per hour. This is about one part in fifteen thousand of the heat generated in the heating coil. In the most unfavourable circumstances it would not have been more than three times this amount.

Water-equivalents.

The water-equivalent of the calorimeter and its accessories was found by carrying out experiments at 20° C., at which temperature the specific heat is, by definition, unity.

From the equation

$$-(\mathrm{M}s+\mathrm{W})\frac{d\theta}{dt}=\frac{\mathrm{E}^2}{\mathrm{RJ}}$$

we have, substituting for $-\dfrac{d\theta}{dt}$ and putting $s=1$,

$$\mathrm{M}+\mathrm{W}=\frac{\mathrm{E}^2}{\mathrm{RJK}\theta^n},$$

in which all the quantities except W are observed.

The value of W at any other temperature was calculated using the temperature-coefficient of the specific heat of copper.

Results.

The specific heat of ordinary water was determined at temperatures between 15° C. and 50° C., the figures obtained being given in Table II., the figures for heavy water in Table III.

TABLE II.

(Water.)

Temperature °C.	Specific heat.	Temperature °C.	Specific heat.
16·85	$1{\cdot}000_7$	38·51	$0{\cdot}998_4$
19·04	$1{\cdot}000_3$	38·81	$0{\cdot}998_5$
19·64	$1{\cdot}000_3$	39·62	$0{\cdot}998_4$
21·52	$0{\cdot}999_8$	39·98	$0{\cdot}998_5$
25·60	$0{\cdot}999_3$	40·87	$0{\cdot}998_6$
25·99	$0{\cdot}999_0$	42·33	$0{\cdot}998_6$
29·80	$0{\cdot}998_8$	43·91	$0{\cdot}998_6$
31·31	$0{\cdot}998_3$	44·97	$0{\cdot}998_7$
31·75	$0{\cdot}998_6$	45·92	$0{\cdot}998_7$
32·89	$0{\cdot}998_4$	46·85	$0{\cdot}998_7$
35·06	$0{\cdot}998_5$	46·90	$0{\cdot}998_9$
36·78	$0{\cdot}998_4$	48·01	$0{\cdot}998_9$
38·02	$0{\cdot}998_4$		

TABLE III.

(Heavy Water.)

Temperature °C.	Specific heat.	Temperature °C.	Specific heat.
11·97	$1{\cdot}009_4$	31·92	$1{\cdot}004_4$
13·88	$1{\cdot}007_9$	32·01	$1{\cdot}004_1$
14·66	$1{\cdot}007_9$	33·83	$1{\cdot}003_8$
14·78	$1{\cdot}007_4$	33·92	$1{\cdot}004_2$
16·72	$1{\cdot}007_5$	35·86	$1{\cdot}004_1$
18·76	$1{\cdot}006_6$	37·63	$1{\cdot}004_2$
19·56	$1{\cdot}006_2$	37·66	$1{\cdot}003_6$
19·75	$1{\cdot}006_1$	38·82	$1{\cdot}003_8$
19·86	$1{\cdot}006_4$	41·00	$1{\cdot}003_3$
21·53	$1{\cdot}005_8$	41·24	$1{\cdot}003_4$
21·75	$1{\cdot}006_1$	42·06	$1{\cdot}003_7$
21·75	$1{\cdot}005_6$	42·35	$1{\cdot}003_1$
23·56	$1{\cdot}005_9$	43·78	$1{\cdot}003_5$
24·69	$1{\cdot}005_1$	47·84	$1{\cdot}004_2$
25·53	$1{\cdot}005_0$	52·65	$1{\cdot}004_9$
30·03	$1{\cdot}004_3$		

The temperature variation of the specific heats of water and of heavy water as determined by these experiments may be represented by equations of the type

$$s = \mathrm{A} + \frac{\mathrm{B}}{\mathrm{C}+t} + \mathrm{D}t + \mathrm{E}t^2 ,$$

where A, B, C, D, and E are constants.

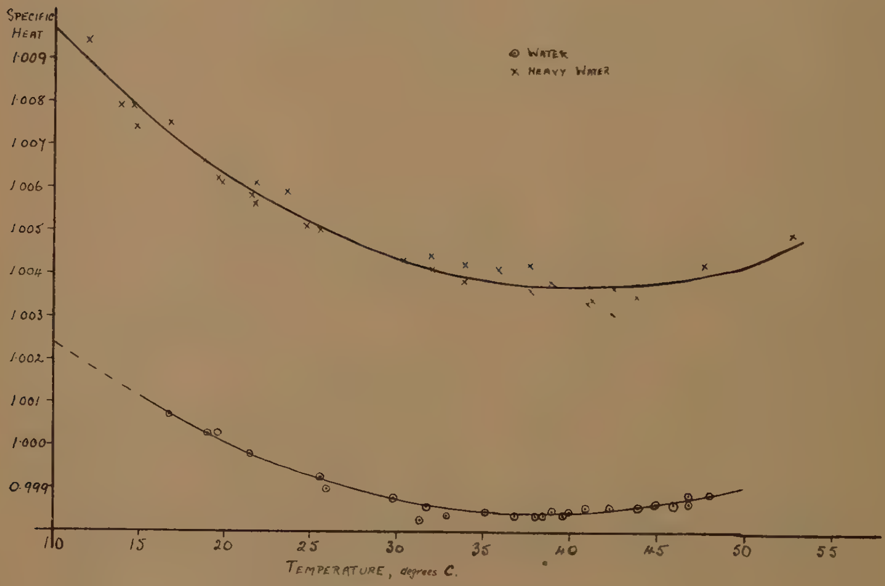
The equation for water is

$$s_{\mathrm{W}} = 1{\cdot}001_8 + \frac{0{\cdot}0974_7}{20+t} - 0{\cdot}000292_6 t + 4{\cdot}22 \times 10^{-6} t^2 ,$$

that for heavy water being

$$s_{\mathrm{H}} = 0{\cdot}995_5 + \frac{0{\cdot}615_9}{30+t} - 0{\cdot}000145_1 t + 3{\cdot}26 \times 10^{-6} t^2 .$$

Fig. 3.



Specific heat—Temperature curves.

These give minimum values of the specific heat at 38·0° C. for water and 41° C. for heavy water. The corresponding specific heats are 0·9984 and 1·0037 respectively. Calculated values of the specific heats at even temperatures between 10° C. and 50° C. are contained in Table IV., columns 3 and 4. Column 2 gives the figures obtained by Jaeger and Steinwehr (2) for ordinary water. In fig. 3 the full lines represent the calculated values, the circles and crosses the experimental figures.

TABLE IV.

Temperature °C.	Water. (J. & S.)	Water. (C. & F.)	Heavy water. (C. & F.)
10	1·0023	$1{\cdot}002_6$	$1{\cdot}009_7$
15	1·0011	$1{\cdot}001_1$	$1{\cdot}007_7$
20	1·0000	$1{\cdot}000_0$	$1{\cdot}006_2$
25	0·9993	$0{\cdot}999_3$	$1{\cdot}005_1$
30	0·9989	$0{\cdot}998_8$	$1{\cdot}004_4$
35	0·9988	$0{\cdot}998_5$	$1{\cdot}003_8$
40	0·9991	$0{\cdot}998_5$	$1{\cdot}003_7$
45	0·9997	$0{\cdot}998_7$	$1{\cdot}003_7$
50	1·0006	$0{\cdot}999_1$	$1{\cdot}004_1$

For heavy water these figures may be compared with those obtained by Brown, Barnes, and Maass (3) :—

B. B. & M.	4–26° C.	1·018	26–45° C.	1·003	26–65° C.	1·008
C. & F.	5–25° C.	$1{\cdot}008_2$	25–45° C.	$1{\cdot}004_2$	25–65° C.	$1{\cdot}004_6$
B. B. & M.	4–65° C.	1·01				
C. & F.	5–65° C.	$1{\cdot}005_9$				

These experiments were carried out in the Physics Department, Queen Mary College, London. Our thanks are due to Prof. H. R. Robinson, who enabled us to have the necessary facilities and also to Imperial Chemical Industries, Ltd., who very generously loaned us the heavy water.

References.

(1) Proc. Phys. Soc. vol. xlv. pt. ii. p. 194 (1933).
(2) *Ann. d. Physik.* vol. lxiv. p. 305 (1921).
(3) ‘Canadian Journal of Research,’ vol. xiii. p. 167 (1935).

XXII. *On the Proton-Proton Scattering.*

By Prof. K. C. KAR, D.Sc., and D. BASU, B.Sc. *.

[Received February 13, 1939.]

THE theory of proton-proton scattering was first developed by Mott † using Temple's ‡ solution for the perturbed wave-equation. In so doing Mott took the perturbation potential to be Coulombian. It has, however, been found subsequently that the experimental value of scattering is much higher than that given by Mott's formula. The theory is therefore extended by Bethe §, and Breit, Condon, and Present ||, who find decidedly better agreement with experiment for certain fitting value of the phase-shift introduced on account of the short range attractive force between the protons.

In the present paper we shall follow the wave statistical method ¶ adopted previously in developing the theory of scattering of electron by atom and of neutron by proton.

Coulomb Potential.

Let us first consider the perturbation potential to be Coulombian. Then, referring the motion to the centre of mass (C-system), we have for the first-order scattering function **

$$\lambda_1\chi_1(r_2, \theta_2) = -\frac{1}{4\pi\sqrt{v}} \cdot \frac{k^2}{\mathrm{E}} \int \mathrm{V}(r_1) e^{ikx_1} \frac{e^{ikr_{12}}}{r_{12}} d\tau_1, \quad . \quad . \quad . \quad (1)$$

where v is the velocity of the proton, $\mathrm{E} = \frac{1}{4}\mathrm{M}v^2$, M being the proton mass and $k = \frac{\pi \mathrm{M}v}{h}$. On proceeding as before, we have

$$\lambda_1\chi_1(r, \theta) = \frac{e^2}{\mathrm{M}v^2} \operatorname{cosec}^2 \frac{\theta}{2} \frac{e^{ikr_0}}{\sqrt{vr}} e^{ik'r_0}, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

* Communicated by Prof. K. C. Kar, D.Sc.

† See Mott and Massey, 'Atomic Collisions,' pp. 75, 76, Eqs. (25), (26).

‡ Temple, Proc. Roy. Soc. A, cxxi. p. 673 (1928).

§ Bethe, Rev. Mod. Phys. viii. p. 131 (1936).

|| Breit, Condon, and Present, Phys. Rev. i. p. 825 (1936).

¶ K. C. Kar, Phil. Mag. xxiv. p. 971 (1937); also Kar and Basu, Phil. Mag. xxvii. p. 76 (1939).

** Kar, Ghosh, and Mukherjee, Phil. Mag. xxiv. p. 964 (1937); also Kar and Basu, *loc. cit.*

where $k'=2k\sin\frac{\theta}{2}$ and r_0, the critical approach, is given by

$$r_0=2\rho\frac{e^2}{Mv^2}\left(\operatorname{cosec}\frac{\theta}{2}+1\right);\quad \rho=1{\cdot}35, \quad . \quad . \quad . \quad . \quad (2.1)$$

in which the additional factor 2 occurs, as the reduced mass is $\frac{1}{2}M$ in the C-system. It may be mentioned that the formula * deduced by Mukherjee and Ghosh is the same as (2) except the last factor, $e^{ik'r_0}$, for which they get $\cos k'r_0$. It is easily seen that they have got the cosine factor because, in the starting integral (1), they have taken $\sin kr_{12}$ for $e^{ikr_{12}}$ in order to make the integrand bounded. This assumption appears to us to be unnecessary. Now in deriving Mukherjee and Ghosh's formula in a rigorous way Kar has made use of Gegenbauer's formula for $\sin kr_{12}$, viz.,

$$\frac{\sin kr_{12}}{r_{12}}=\pi\sum_s(s+\tfrac{1}{2})\frac{J_{s+\frac{1}{2}}(kr_1)}{\sqrt{r_1}}\cdot\frac{J_{s+\frac{1}{2}}(kr_2)}{\sqrt{r_2}}P_s(\cos\theta_1)P_s(\cos\theta_2). \quad . \quad (3)$$

It is thus evident that in deriving (2) we have to use, in addition, Gagenbauer's cosine formula, viz.,

$$\frac{\cos kr_{12}}{r_{12}}=\pi\sum_s(-1)^s(s+\tfrac{1}{2})\frac{J_{-s-\frac{1}{2}}(kr_1)}{\sqrt{r_1}}\cdot\frac{J_{s+\frac{1}{2}}(kr_2)}{\sqrt{r_2}}P_s(\cos\theta_1)P_s(\cos\theta_2). \quad (4)$$

The method being otherwise the same as that given by Kar, the successive steps leading to (2) are not given.

Now, on substituting for k' and r_0, we have from (2),

$$\lambda_1\chi_1(r,\theta)=\frac{e^2}{Mv^2}\operatorname{cosec}^2\frac{\theta}{2}\frac{e^{ikr}}{\sqrt{vr}}e^{2i\rho\alpha(\left(1+\sin\frac{\theta}{2}\right)}, \quad . \quad . \quad . \quad . \quad (5)$$

where

$$\alpha=\frac{2\pi e^2}{hv} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.1)$$

In laboratory coordinates (L-system), Eq. (5) becomes

$$\lambda_1\chi_1(r,2\theta)=\frac{e^2}{Mv^2}\operatorname{cosec}^2\theta\frac{e^{ikr}}{\sqrt{vr}}e^{2i\rho\alpha(1+\sin\theta)}. \quad . \quad . \quad . \quad . \quad (6)$$

The incident and bombarded proton cannot be distinguished experimentally; therefore we have to calculate the total number for both. The scattering function in L-system for the bombarded proton is evidently

$$\lambda_1\chi_1(r,\pi-2\theta)=\frac{e^2}{Mv^2}\sec^2\theta\frac{e^{ikr}}{\sqrt{vr}}e^{2i\rho\alpha(1+\cos\theta)}. \quad . \quad . \quad . \quad . \quad (7)$$

---

* *Vide* Kar, Phil. Mag. *loc. cit.*; see also K. K. Mukherjee, *Phys. Zeit.* xxxiv. p. 175 (1933), and M. Ghosh, Phil. Mag. xx. p. 234 (1935).

The total scattering function should be taken as the sum of (6) and (7) if the functions are symmetrical, *i. e.*, if the proton spins are anti-parallel. On the other hand we should take their difference if the functions (6) and (7) are anti-symmetrical, *i. e.*, the spins are parallel. In the former case we have for the relative intensity of scattering per unit solid angle, since $2\pi \sin 2\theta \, d(2\theta) = 4 \cos\theta \, . \, 2\pi \sin\theta \, d\theta$,

$$I_1 = \left(\frac{e^2}{Mv^2}\right)^2 [\operatorname{cosec}^4\theta + \sec^4\theta$$
$$+ 2 \operatorname{cosec}^2\theta \sec^2\theta \cos\{2\rho\alpha(\sin\theta - \cos\theta)\}] 4 \cos\theta . \quad . \quad (8)$$

Similarly, in the latter case,

$$I_2 = \left(\frac{e^2}{Mv^2}\right)^2 [\operatorname{cosec}^4\theta + \sec^4\theta$$
$$- 2 \operatorname{cosec}^2\theta \sec^2\theta \cos\{2\rho\alpha(\sin\theta - \cos\theta)\}] 4 \cos\theta . \quad . \quad (9)$$

As the weights of the two classes of events are as 1 : 3, we have for the total intensity of scattering

$$I = \tfrac{1}{4} I_1 + \tfrac{3}{4} I_2$$
$$= \left(\frac{e^2}{Mv^2}\right)^2 [\operatorname{cosec}^4\theta + \sec^4\theta$$
$$- \operatorname{cosec}^2\theta \sec^2\theta \cos\{2\rho\alpha(\sin\theta - \cos\theta)\}] 4 \cos\theta, \quad . \quad . \quad (10)$$

which is different from Mott's formula. It may, however, be noted that since $\alpha \ll 1$ and so $\cos\{\ \} \rightarrow 1$, (10) reduces to

$$I = \left(\frac{e^2}{Mv^2}\right)^2 [\operatorname{cosec}^4\theta + \sec^4\theta - \operatorname{cosec}^2\theta \sec^2\theta], \quad . \quad . \quad (10.1)$$

to which also Mott's formula approximates. We shall therefore refer to (10.1) hereafter as the MKB-formula. It does not agree with experiment as mentioned before.

The Effect of the Attractive Force.

It is evident that between protons there should exist a short-range attractive force of the nature of neutron-proton attraction. The effect of such a force is evidently to modify the scattering formula obtained on the assumption of Coulomb repulsion. Now it has been shown * already that for neutron-proton attraction the relative intensity of scattering is symmetrical and is independent of the incident velocity. We have, then, for the modified scattering function (*vide* (6))

$$\lambda_1 \chi_1(r, 2\theta) = \frac{e^2}{Mv^2} \operatorname{cosec}^2\theta \frac{e^{ikr}}{\sqrt{vr}} e^{2i\rho\alpha(1+\sin\theta)} - g \frac{e^{ikr}}{\sqrt{vr}}, \quad . \quad . \quad (11)$$

---

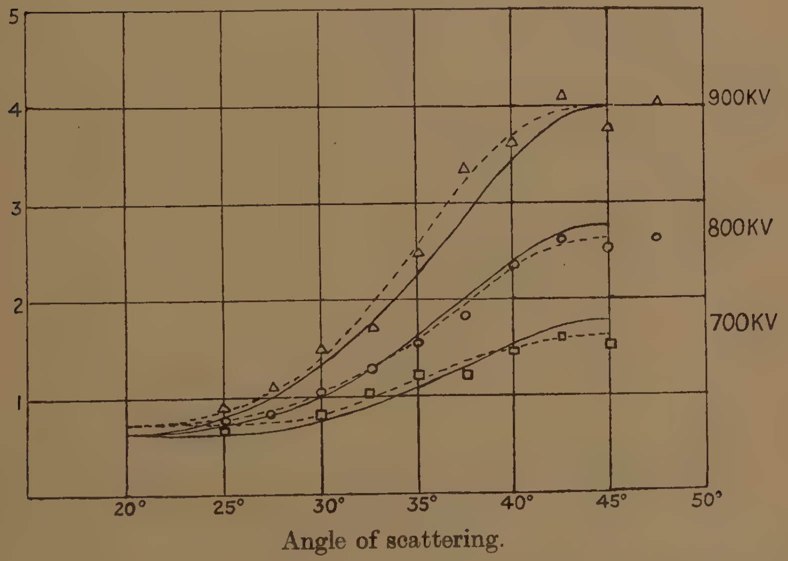
* Kar and Basu, *loc. cit.*

and for exchange (*vide* (7))

$$\lambda_1\chi_1(r,\pi-2\theta)=\frac{e^2}{\mathrm{M}v^2}\sec^2\theta\,\frac{e^{ikr}}{\sqrt{v}\,r}\,e^{2i\rho\alpha(1+\cos\theta)}-g\,\frac{e^{ikr}}{\sqrt{v}\,r}, \quad . \quad . \quad (12)$$

where g is independent of v and the angle of scattering. It is taken with a minus sign as the force is attractive, whereas the former Coulomb force is repulsive. On proceeding as before we have for unlike spin

$$\mathrm{I}_1=\left(\frac{e^2}{\mathrm{M}v^2}\right)^2\Big[\operatorname{cosec}^4\theta+\sec^4\theta+2\operatorname{cosec}^2\theta\sec^2\theta\cos\{2\rho\alpha(\sin\theta-\cos\theta)\}$$
$$-4\,\frac{\mathrm{M}v^2}{e^2}\,g\operatorname{cosec}^2\theta\cos\{2\rho\alpha(1+\sin\theta)\}$$
$$-4\,\frac{\mathrm{M}v^2}{e^2}\,g\sec^2\theta\cos\{2\rho\alpha(1+\cos\theta)\}+4\left(\frac{\mathrm{M}v^2}{e^2}\right)^2g^2\Big]\,4\cos\theta, \quad (13)$$



but for like spin I_2 is unmodified and is given by (9). Therefore we have for the total intensity corresponding to (10)

$$\mathrm{I}=\left(\frac{e^2}{\mathrm{M}v^2}\right)^2\Big[\operatorname{cosec}^4\theta+\sec^4\theta-\operatorname{cosec}^2\theta\sec^2\theta\cos\{2\rho\alpha(\sin\theta-\cos\theta)\}$$
$$-\frac{\mathrm{M}v^2}{e^2}\,g\operatorname{cosec}^2\theta\cos\{2\rho\alpha(1+\sin\theta)\}$$
$$-\frac{\mathrm{M}v^2}{e^2}\,g\sec^2\theta\cos\{2\rho\alpha(1+\cos\theta)\}+\left(\frac{\mathrm{M}v^2}{e^2}\right)^2g^2\Big]. \quad . \quad . \quad . \quad (14)$$

For energies near 1MV, $\alpha \ll 1$, therefore $\cos\{\ \} \rightarrow 1$, and so (14) becomes

$$\mathrm{I} = \left(\frac{e^2}{\mathrm{M}v^2}\right)^2 \left[\operatorname{cosec}^4\theta + \sec^4\theta - \operatorname{cosec}^2\theta \sec^2\theta \left(1 + \frac{\mathrm{M}v^2}{e^2} g\right) + \left(\frac{\mathrm{M}v^2}{e^2}\right)^2 g^2\right]. \quad . \quad . \quad . \quad (14.1)$$

Taking the ratio of the above theoretical value of the intensity of scattering and the value given in (10.1) we have

$$\frac{\mathrm{I}}{\mathrm{I}_{\mathrm{KBM}}} = \frac{\operatorname{cosec}^4\theta + \sec^4\theta - \operatorname{cosec}^2\theta \sec^2\theta \left(1 + \frac{\mathrm{M}v^2}{e^2} g\right) + \left(\frac{\mathrm{M}v^2}{e^2}\right)^2 g^2}{\operatorname{cosec}^4\theta + \sec^4\theta - \operatorname{cosec}^2\theta \sec^2\theta} \quad . \quad (15)$$

The experimental value of the ratio (15) is determined by Tuve, Heydenburg and Hafstad * for different incident velocities. The theoretical and experimental values are, however, compared in the above figure only for high velocities, viz., 700 kv., 800 kv., and 900 kv., for which the departure from MKB-formula is expected to be appreciable. In calculating the theoretical values we have taken $g = 4{\cdot}766 \times 10^{-13}$, which is its fitting value at 45° for 900 kv. It is evident from the above figure, where the theoretical curves are drawn continuous, that the agreement is quite close for all velocities. This appears to be rather significant, as in the previous theory given by Breit and others (*loc. cit.*) the fitting value of the phase shift is taken differently for different incident velocities.

Physical Laboratory,
Presidency College, Calcutta, India.
January 1939.

---

* Tuve, Heydenburg, and Hafstad, Phys. Rev. l. p. 806 (1936).

XXIII. *Secrecy by Inversion of Syllables.*

By Mario Marro, Dr. Ing.*

[Received March 16, 1939.]

Following my experimental researches on the dissecting of speech at regular intervals of time †, I investigated what would happen when the syllables were periodically inverted.

The results are shown below :—

Number of inversions per second.	Observed results.
100	Voice is comprehensible, but it is as if the speaker suffered from a cold.
200	Voice is comprehensible, but becomes more altered.
300 500	The deformation becomes greater.
600	Voice is comprehensible with difficulty.
700	Voice is practically incomprehensible.
800 900	Voice is incomprehensible, but at the same time one hears a whispering which allows some words to be intelligible.
1000	Voice is quite incomprehensible.

The importance of the frequency of inversions is justified, as the voice comprises a great number of harmonics, and the deformation of speech depends naturally on the inversions of the said harmonics.

As an inverter I used the arrangement represented in fig. 1. Two tubes A and B act as diodes. The voltage to the plates is supplied by the voltage at the terminals of the secondary winding of the microphone transformer T when one speaks before the microphone M. The grids of tubes A and B are given positive and negative voltages alternately, supplied by the oscillator O. The frequency of this oscillation is determined by the variable condenser C.

During the time when the grid of valve A becomes positive, the grid of valve B becomes negative, and conversely. Thus when the grid of valve A is positive the voltage supplied to the plate of the valve by that at the secondary winding of the microphone transformer, creates a current in the listening line L following the direction of the arrow a_1. During this time no current flows through the valve B, because its grid is negative.

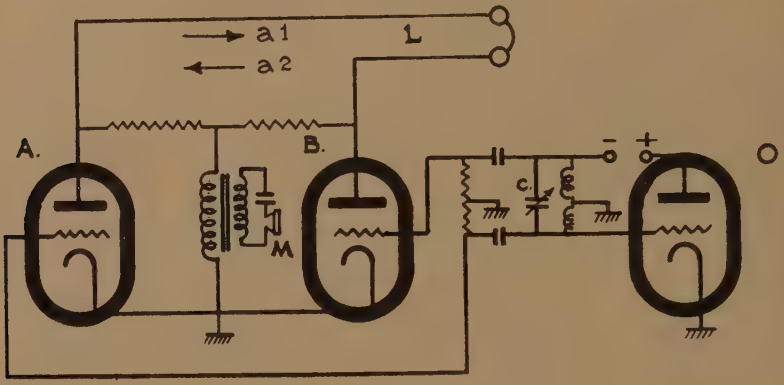
---

* Communicated by the Author.

† "On the Persistence of the Sensation of Speech," Phil. Mag. xxii. p. 847 (1936) ; "Two-way Speech by Wireless," Phil. Mag. xxviii. p. 248 (1939).

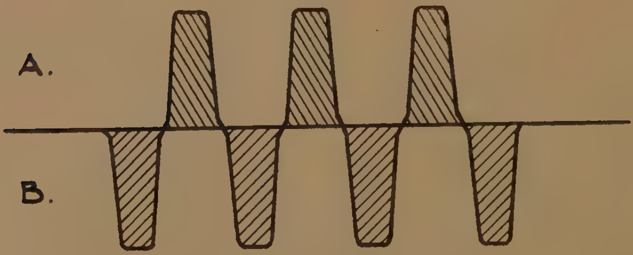
During the succeeding time, when the grid of valve B is positive, the current in the listening line follows the direction of the arrow a_2. Thus the syllables of the speech are periodically inverted with a frequency equal to that generated by the oscillating valve O.

Fig. 1.



Inverter arrangement.

Fig. 2.



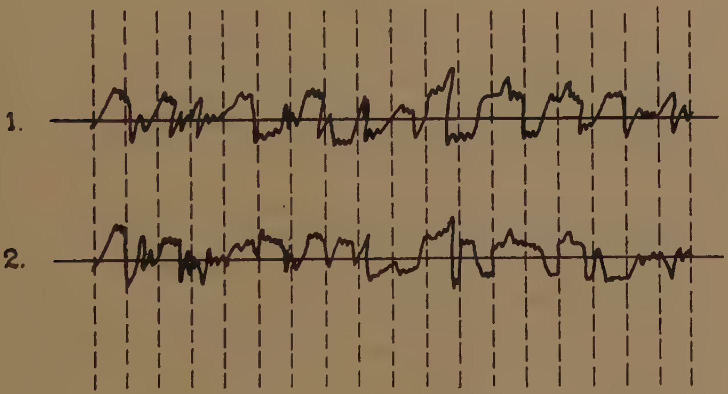
Inversion times.

One will then hear an unknown language. When one does not speak before the microphone no current flows through the listening line.

If the voltage generated by the oscillator O is high enough so that the valves A and B will work near their saturation point, the inversion times are as those represented in fig. 2. With only a valve A or B, the voice is fully intelligible, as in these circumstances the speech will be transmitted at regular intervals of time, and it will be the case indicated in my note "Two-way Speech by Wireless."

Several secret systems are actually used in wireless telephony. Generally the frequencies of speech extending from 250 to 2750 cycles are used to modulate a 3000 cycles note; then two sidebands, 250 to 2750 cycles and 3250 to 5750 cycles, are formed. Selecting the lower sideband, it will be observed that the speech frequencies are transposed, for it is the 2750 cycles component of speech which produces the 250 cycles frequency, whereas the 250 cycles component produces the 2750 cycles frequency. But the simple method of transposition indicated above is not adopted because of the difficulty of separating the transposed speech from the original speech and from the 3000 carrier. Instead a double modulation system is used. But it is necessary to employ a number of filters for separating the various sidebands and the system becomes very complex. The object of my research was to find a secret system, very simple, of little weight and small size, suitable to be carried on an aeroplane.

Fig. 3.



1 = Speech Current. 2. = The same Inverted.

Speech inversion.

The arrangement which I studied is reversible. That is, when the inverted speech enters into an arrangement such as that described, it will come out perfectly intelligible. It is necessary to have perfect synchronism between the two correspondent systems, that is, between the frequencies of the two oscillators O. These frequencies may be controlled by a fork.

Fig. 3 shows how the inversion of speech takes place. In these circumstances we cannot speak of modulation of the inversion frequency by the characteristic frequencies of the voice.

Via Indipendenza, 59,
Bologna, Italy.

XXIV. *Notices respecting New Books.*

The Raman Effect and its Chemical Applications. By JAMES H. HIBBEN. [Pp. 544.] (New York: Reinhold Publishing Corporation. London: Chapman & Hall. 1939. Price 66*s.* net.)

IT is difficult to believe that a space of but eleven years divides us from the time when Sir C. V. Raman gave an address on " A New Radiation," in which he described those experiments on scattering and fluorescence which led to an explanation in terms of what is most rightly described as the Raman effect. The immense mass of literature that has accumulated round the subject since 1928 might well be assumed to be the result of a generation's work.

All the more necessary is it that an ordered and critical account of this work should be undertaken, and the task has been carried out with conspicuous success by Mr. Hibben. The massive volume under review is far more than a summary—it may be fittingly described as an encyclopædia of the Raman effect. Very few aspects of the effect are neglected by the learned author, and his exposition is both clear and accurate. The specialist will welcome the book; and the physicist who wishes to have something more than a perfunctory knowledge of the effect will find the book absorbing, interesting, and exactly suited to the needs of one who wishes to make a start on experimental work; for Mr. Hibben devotes sufficient space to a description of experimental technique to assist materially anyone desiring to begin research in this field.

It is impossible even to give the heads of the twenty-nine chapters of the book, but it will sufficiently indicate its scope if it is pointed out that the work consists of three main sections—a general discussion of the Raman effect, the Raman spectra of organic compounds, and the Raman spectra of inorganic compounds.

The book is provided with a bibliography of nearly two thousand entries, the value of which is considerably enhanced by the addition of an index of compounds, which points to the references in the bibliography containing accounts of the work done on any particular compound.

The volume may be fairly described as indispensable to anyone working on the Raman effect, and the author is to be heartily congratulated on a notable addition to scientific literature.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XXV. *Stability and Configuration of the Wakes produced by Solid Bodies moving through Fluids.*

By S. H. Hollingdale, M.A., Ph.D., D.I.C.*

[Received February 2, 1939.]

[Plates IX.–XIII.]

Contents.

Page
Part I.—The Stability of Flow in an Unlimited Viscous Fluid .. 211
1. Introduction 211
2. The Governing Equations of the Disturbances ... 212
3. The Existence of Neutral Oscillations 213
4. The Types of Profile 214
  4.1. The Symmetrical Profile 214
  4.2. The Monotonic Profile 215
5. The Wave-velocity of the Oscillation 215
. Application to Particular Velocity Profiles 217
  6.1. Method of Approximation to the Velocity Profile 217
  6.2. The Particular Profiles 218
  6.3. Definition of the Reynolds Number of the Profiles 218
  6.4. The Velocity Profile $U(y)=\exp\left(\frac{1-y^2}{2}\right)$.. 221
  6.5. The Velocity Profile $U(y)=\operatorname{erf} y$ 224
7. Symmetrical Oscillations 225
8. Summary 226

Part II.—Experimental Investigations of the Wake and Comparison with Theory 227
9. Introduction 227
10. Apparatus and Experimental Method........... 228
11. Experiments Performed and General Remarks .. 230
12. Notation..................................... 233
13. Comparison with Theory 234
14. The Velocity Field in the Laminar Wake behind a Flat Plate................................. 235
  14.1. Velocity along Line of Symmetry 236
  14.2. Velocity Distribution along Lines at Right Angles to the Line of Symmetry....... 237

---

* Communicated by Prof. L. Bairstow.

Page
15. Experimental Calculation of the "Drag" of the Flat Plate 239
16. Velocity Field in the Wake behind an Aerofoil Section 242
    16.1. Laminar Wake. R=631 242
    16.2. Oscillating Wake. R=1377 245
17. Comparison of Theory and Experiment for the Oscillating Wake. 247
18. The Observed Oscillating Wake 248
19. Correlation of Theoretical and Experimental Units. 249
20. Measurement of the Representative Velocity and Length 250
21. Difficulties of the Comparison 251
22. Comparison for Wakes produced by a Flat Plate . 252
    22.1. F. 17–20. R=791 252
    22.2. F. 21–24. R=1110 253
    22.3. F. 9–12. R=1090 254
23. Comparison for Wake produced by an Aerofoil Section 254
24. The Influence of the Tank Walls 254
25. Discussion of Results 256
26. Summary 257

TABLES.

I. Summary of Oscillation Calculations for Laminar Profiles. 226
II. Particulars of Plates measured to give Velocity Distribution in the Wake 233
III. Summary of Area and Drag Calculations for Laminar Flat Plate Wakes 241
IV. Summary of Area and Drag Calculations for Laminar Aerofoil Wake 243
V. Effect of Choice of Origin on Theoretical Velocity Field . . 244
VI. Summary of Area and Drag Calculations for Oscillating Aerofoil Wake 247
VII. Particulars of Measured Oscillating Wakes 249
VIII. Accuracy of Approximation used in deriving the Representative Length of the Profile 251
IX. Comparison of Theory and Experiment for Flat Plate Wake R=791 253
X. Comparison of Theory and Experiment for Flat Plate Wake R=1110 253
XI. Comparison of Theory and Experiment for Aerofoil Wake R=1377 254
XII. Comparison allowing for Effect of the Walls of the Tank . . 255

ABSTRACT.

Part I.

The stability of certain types of "one-dimensional" steady flow in an unlimited viscous fluid has been investigated by the method of "small oscillations." The general result is obtained that the velocity of the oscillations (in the direction of motion of the main flow) is equal to the velocity of the main flow at the point of inflexion of the laminar velocity profile.

Two profiles of physical interest are considered in greater detail and approximate values obtained for the wave-length of the oscillations.

Part II.

Photographs have been obtained of the wakes produced by models (flat plates and symmetrical areofoil sections at zero incidence) moving in a tank of water. Oscillations do not occur when the Reynolds number is less than 600 for the flat plate and 1000 for the aerofoil section, and it is probable that the onset of instability can be considerably delayed by careful elimination of all "residual motions" of the water.

The velocity field in the laminar wake has been measured, and a comparison made with certain results of Goldstein and Fage. Good agreement is obtained, the divergence being appreciable at the edges of the wake, due to the effect of the walls of the tank.

Measurements of the wave-length and wave-velocity have been made for various oscillating wakes, for a comparison with the conclusions of Part I. The agreement between theory and experiment is, on the whole, satisfactory, but the comparison is somewhat blunted owing to several difficulties in the interpretation of the theoretical results.

PART I.—THE STABILITY OF FLOW IN AN UNLIMITED VISCOUS FLUID.

1. *Introduction.*

THE problem of the stability of fluid flow has engaged the attention of a large number of investigators since Stokes*, in 1843, first suggested the possibility of unstable solutions of the equations of motion. Previous investigations have been mainly concerned with fields of flow which are delimited by solid boundaries; in such cases the existence of "boundary layers" within which viscosity effects cannot be neglected is of great importance in the determination of the stability of the flow. In this paper, on the other hand, we are concerned with certain types of flow in an unlimited fluid.

---

* Camb. Phil. Soc. Trans. viii. p. 105 (1843).

The greatest advance in this difficult subject has been made by Tollmien in a recent paper *, in which he derived a general criterion for the instability of laminar velocity distributions. Tollmien considered two-dimensional velocity profiles in which the main velocity distribution is dependent only on the coordinate normal to the direction of flow. On the basis of a very general treatment he obtained the simple result that velocity profiles having a point of inflexion are unstable. The effect of friction was assumed to be negligible and did not enter into his discussion of stability, and so his results, while explicitly applied in his paper to two important classes of flow bounded by solid walls, appear equally applicable to unlimited fields of flow. Tollmien also proceeded a stage further and derived some characteristics of oscillations of increasing amplitude.

The experimental work described in Part II. of this paper was carried out between September 1933 and March 1935. The theoretical treatment was developed during the same period in ignorance of Tollmien's work (T. I.), and is here presented as a simpler, although less rigorous, treatment than Tollmien's †, and also to afford a comparison with the experimental results which are given in Part II. of this paper.

2. *The Governing Equation of the Disturbance.*

The fluid is assumed to be viscous, incompressible, and unlimited in extent. The velocity of the main steady flow, on which the disturbance is to be superposed, is assumed to depend only on the coordinate y at right angles to the direction of flow and, with an obvious choice of axes and notation, may be represented by

$$u=\mathrm{U}(y), \quad v=0, \quad w=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Velocity distributions which are approximately of this type are frequently encountered, an example being the "wake" formed behind a symmetrical streamline cylinder moving through a fluid in the direction of its axis of symmetry. The assumption of equation (1) would then imply neglect of the broadening of the wake down stream. (The photographs of such wakes in Part II. of this paper show that this broadening is very small.)

Attention will be confined to two-dimensional disturbances, as an investigation by Squire ‡ indicates that no limitation of the problem is thereby introduced. Following the usual procedure, we shall superpose a small oscillatory disturbance on the main flow. A non-dimensional system of units may be employed by the introduction of a representative length, h, and a representative velocity, U_m. The fundamental equation governing the development of the disturbance may be written in the standard form,

---

* *Nachrichten Ges. Wiss. Göttingen, Math.-Phys. Klasse*, p. 79 (1935). (Referred to subsequently as T. I.)

† The mode of treatment here adopted is, in fact, indicated by Tollmien as a possible alternative.

‡ Proc. Roy. Soc. A, cxlii. p. 621 (1933).

$$(U-c)(\phi''-\alpha^2\phi)-U''\phi=-\frac{i}{\alpha R_m}(\phi''''-2\alpha^2\phi''+\alpha^4), \quad . \quad . \quad . \quad (2)$$

where differentiations are made with respect to the non-dimensional y-coordinate and the characteristic Reynolds number is given by $R_m=U_m h/\nu$. The linear form of the differential equation (2) enables us to build up the disturbance stream function, ψ, from partial vibrations which are assumed to be of the form *

$$\psi=\phi\,(y)\,.\exp i\alpha(x-ct), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where α is a real constant, but c is, in general, complex. The wave-length, λ, of the disturbance is, of course, given by $\lambda=2\pi/\alpha$. Writing $c=c_r+ic_i$, where c_r and c_i are real, we see that c_r is the "wave-velocity" of the disturbance and (αc_i) is the "logarithmic increment" and is positive for vibrations of increasing amplitude and negative for damped vibrations.

3. *The Existence of Neutral Oscillations.*

We shall now assume c to be entirely real, *i. e.*, consider oscillations which are neither increasing in amplitude nor are damped. The existence of real solutions in this case will then indicate instability for sufficiently large values of the Reynolds number.

Following previous workers in the subject †, we shall consider the case of large R_m and discuss the validity of neglecting the viscosity terms in equation (2). The equation then reduces to the disturbance equation for an inviscid fluid,

$$(U-c)(\phi''-\alpha^2\phi)-U''\phi=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

In general a singular point of this equation occurs at the point (or points) for which $U=c$. A consideration of the analytical properties of the integrals of this equation shows that an infinity occurs in the gradient of ϕ at this "critical point", and the effects of friction must be taken into account even for very large values of R_m. Tollmien ‡ has shown that transit through this critical point introduces a phase discontinuity into the distance, represented mathematically by the appearance of complex values of ϕ.

---

* In physical applications the real part of ψ will be taken.

† In particular, reference may be made to Tollmien, *Nachrichten Ges. Wiss. Göttingen, Math.-Phys. Klasse*, p. 21 (1929). (Referred to subsequently as T. II.)

‡ See T. I. and T. II.

Solberg * has shown that $(U-c)$ vanishes for some point of the field for a symmetrical profile, and in his recent paper † Tollmien has extended the proof for all profiles of the types here discussed.

We shall now explore the consequences of the hypothesis that $U=c$ when $U''\neq 0$.

4. *The Types of Profile.*

So far the only limitation that has been imposed on the velocity profile is that it depends on only one space coordinate. We shall further assume that $U(y)$ tends to a constant value as $y\rightarrow\pm\infty$; profiles with any type of back-flow are also excluded. We shall limit the discussion to profiles of two types :—

(1) Symmetrical profiles. Hence, by a suitable choice of axes $U(y)$ may be taken as an even function of y. There will be two critical points given by $y=\pm y_c$, where $y_c>0$.

(2) Profiles for which $U(y)$ is a monotonic function of y, with only one point of inflexion which, by suitable choice of axes, may be taken at the origin. There will also only be one critical point $(U=c)$, which may be written as $y=y_c$.

These two types include most profiles of physical interest, and we shall consider one profile of each type in greater detail.

4.1. *The Symmetrical Profile.*

In this case any function satisfying equations (2) or (4) may be separated into two parts, respectively odd and even in y, each of which separately must be a solution of the differential equation. The even solution corresponds to unsymmetrical oscillations, the odd solution to symmetrical oscillations. We may consider these two types separately.

Thus, near the axis of symmetry, $y=0$, we shall have a series solution of equation (4), either odd or even in y, of the forms

$$\phi\equiv\phi_{\mathrm{I}}=a_0+a_1y^2+a_2y^4+\ldots \qquad (5\,a)$$

or

$$\phi_{\mathrm{I}}=b_0y+b_1y^3+b_2y^5+\ldots \qquad (5\,b)$$

This symmetry condition allows us to consider only positive values of y. Since the coefficients of the differential equation (4) are real, and the two solutions (5 *a*) and (5 *b*) are independent, they may be expressed as real functions, except for possible multiplying constants. In the range $0<y<y_c$ the general solution of equation (4) may be written in the form

---

* Proc. of Internat. Congress of Applied Mechanics. Delft, 1924.

† T. I.

$$\left.\begin{aligned}\phi&=A\phi_1+B\phi_2\\&\equiv A\phi_1+B[\phi_{21}+\phi_1\log(y_c-y)],\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where ϕ_1 is the solution of equation (4) that is regular at the critical point $y=y_c$, A and B are constants, and ϕ_1 and ϕ_2 have been chosen to be real. The function ϕ_{21} will also be regular at the critical point, but ϕ_2 is of the form

$$\phi=\phi_1\int^y\frac{dy}{\phi_1^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

The constants A and B must be chosen so that ϕ joins on to ϕ_{I} in the range $0<y<y_c$, from which it follows that the ratio A/B is real.

4.2. *The Monotonic Profile.*

When $U(y)$ is a monotonic function of y, the condition at $y\rightarrow-\infty$ replaces that at $y=0$. Since the profile tends to the straight line U=constant as $y\rightarrow-\infty$ and ϕ must tend to zero, the solution of equation (4) tends to

$$\phi_0=K.e^{+\alpha y} \text{ as } y\rightarrow-\infty, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where K may, in general, be complex.

Equations (6) and (8) now represent the solution of the differential equation in the range $-\infty<y<y_c$. A and B are now determined by the condition that ϕ must "join on" to ϕ_0 in this range, and again A/B must necessarily be real.

5. *The Wave-velocity of the Oscillation.*

The solution of equation (4) is now completely determined for both types of profile when $y<y_c$ (or, more strictly, for the symmetrical profile, when $|y|<y_c$). We now consider transit through the critical point involving a logarithmic infinity in ϕ_2'.

When $y>y_c$ we can write,

$$\log(y_c-y)=\log|y_c-y|\pm i\pi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

but have no means of deciding which sign should be taken. The whole problem is thus rendered indeterminate, and it is necessary to introduce viscosity in order to remove the ambiguity of sign. Tollmien * has provided a mathematical discussion of the complex integrals involved in the transition, and has shown that the negative sign must be taken. The whole discussion of this transit through the critical point has made clear the inadmissibility of neglecting viscosity effects, as was done by Rayleigh and most early workers in the subject.

---

* T. II.

We can now write the solution for $y>y_c$ as

$$\phi=A\phi_1+B[\phi_{21}+\phi_1(\log|y_c-y|-i\pi)]. \quad . \quad . \quad . \quad (10)$$

As $y\rightarrow+\infty$, the profile tends to the straight line U=constant, and the solution of equation (4) tends to

$$\phi=C.e^{-\alpha y}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where C may, in general, be complex.

Hence the expression (10) for ϕ must tend to the expression (11) when $y\rightarrow+\infty$.

We shall now show that this is not possible, in view of the assumption that we have made as to the non-vanishing of U'' at the critical point.

When $\alpha=0$, equation (4) reduces to

$$(U-c)\phi''-U''\phi=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

of which $\phi=(U-c)$ is a particular solution. The general value of ϕ_1 may be assumed to be capable of expansion in powers of α^2, thus

$$\phi_1=(U-c)+\alpha^2f_1+\alpha^4f_2+\ldots+\alpha^{2n}f_n+\ldots \quad . \quad . \quad . \quad (13)$$

where f_n is a function of y, c, and U.

Since α^2 is independent of y, on substituting in equation (7) we obtain an identity in which coefficients of powers of α^2 may be separately equated to zero.

Writing

$$f_n=(U-c)\gamma_n,$$

we obtain

$$(U-c)\gamma_1''+2U'\gamma_1'=(U-c),$$

giving

$$\gamma_1=\int_{y_c}^{y}\frac{1}{(U-c)^2}\int_{y_c}^{y_2}(U-c)^2\,dy_1\,dy_2. \quad . \quad . \quad . \quad . \quad . \quad (14)$$

For the general term we have

$$(U-c)\gamma_n''+2U'\gamma_n'=(U-c)\gamma_{n-1},$$

giving

$$\gamma_n=\int_{y_c}^{y}\frac{1}{(U-c)^2}\int_{y_c}^{y}(U-c)^2\,\gamma_{n-1}\,dy_1\,dy_2. \quad . \quad . \quad . \quad . \quad (15)$$

The lower limit of the integrals may be chosen arbitrarily in the range (y_c, ∞); y_c has been chosen for simplicity. The regularity of f_n near the critical point is apparent by noting that near y_c, $(U-c)$ tends to $(U')_{y=y_c}.(y-y_c)$.

Referring to equation (14), since $(U-c)^2$ is always positive, it follows that γ_1 is always positive when $y>y_c$. It is clear from equation (15) that the general term γ_n is likewise positive in this range.

In general, therefore, ϕ_1 will not tend to zero as $y \to +\infty$. It may do so for certain characteristic values of α^2, but since α is independent of y, and all possible values of $\alpha(>0)$ are being considered, this possibility will not affect the general argument.

It follows that the expression for ϕ given by equation (10) cannot tend to the expression given by equation (11) when $y \to +\infty$.

This contradiction implies that the initial supposition is incorrect, and we are in a position to state the basic result, namely,

$$(\mathrm{U}-c)=0 \quad \text{when} \quad \mathrm{U}''=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (16)$$

or the wave-velocity, c, is equal to the velocity of the fluid at the point of inflexion of the velocity profile. The singular point is thus removed from the solution of equation (4), and we are enabled to approximate to the complete equation (2) by means of equation (4) throughout the whole range.

6. *Application to Particular Velocity Profiles.*

We will now apply the above proposition to a consideration of two particular profiles. From the nature of the treatment we see that no critical Reynolds number for the onset of instability will be obtained.

The velocity profile only occurs in equation (4) in the form $\mathrm{U}''/(\mathrm{U}-c)$. For the symmetrical profile this expression is clearly an even function of y. For the " monotonic profile " c will be zero, since the axes have been chosen so that $\mathrm{U}''(y)=0$ at the origin. Hence in this case, also, $\mathrm{U}''/(\mathrm{U}-c)$ will be an even function.

The complete solution of equation (4) may, therefore, in both cases, be separated into two parts, each of which separately satisfies the differential equation. It is probable from physical considerations that symmetrical disturbances (corresponding to ϕ an odd function of y) will not produce instability. This will be illustrated later (§ 7) with reference to particular examples, and we shall meanwhile confine attention to solutions for which ϕ is an even function of y.

6.1. *Method of Approximation to the Velocity Profile.*

Profiles of the type $\mathrm{U}(y)=\mathrm{A}\cos(ny+\epsilon)$ and $\mathrm{B}\cosh(my+\epsilon')$ allow the equation (4) to be solved exactly in finite terms. This suggested the possibility of approximating to the given profile by linear combinations of terms of trigonometrical or hyperbolic type. $\mathrm{U}(y)$ and $\mathrm{U}'(y)$ are made continuous throughout the whole range, and arbitrary constants in the resulting partial solutions for ϕ are chosen so that ϕ and ϕ' are always continuous. On eliminating the constants of integration involved, a period equation is obtained for α.

6.2. *The Particular Profiles.*

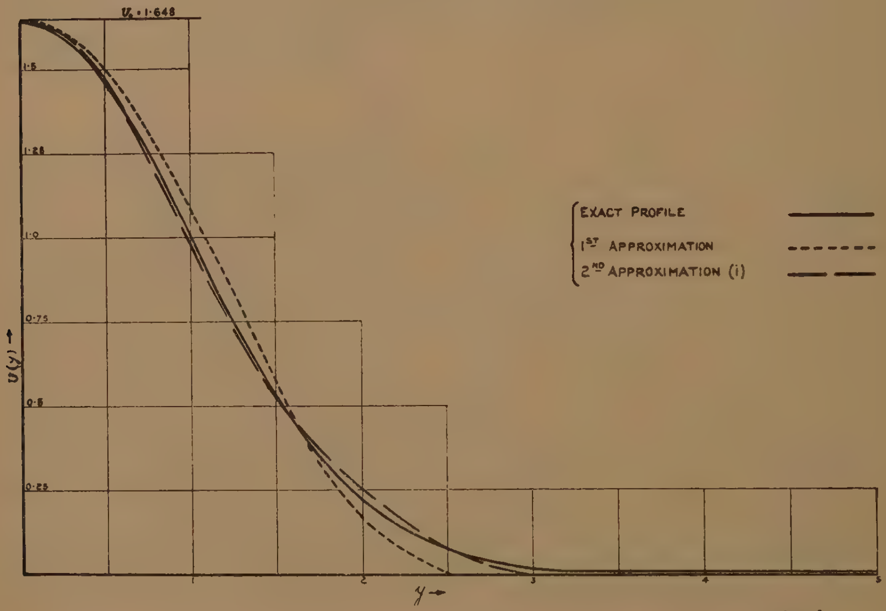
The result expressed in equation (16) will now be applied to investigate the stability of one profile of each of the two types previously specified:—

$$(1) \qquad U(y)=e^{\frac{1-y^2}{2}} \quad \dots\dots\dots\dots \quad (17)$$

$$(2) \qquad U(y)=\frac{2}{\sqrt{\pi}}\int_0^y e^{-y^2}\,dy \equiv \operatorname{erf}(y) \quad \dots\dots \quad (18)$$

The former is a symmetrical profile, with points of inflexion at $y=\pm 1$. A profile of this form provides a close approximation to the flow in the "wake" produced behind a "smooth" body moving through a fluid, and so provides suitable results for comparison with experiment.

Fig. 1.



Comparison of exact and approximate velocity profiles $U(y)=e^{\frac{1-y^2}{2}}$.

The latter is a "monotonic" profile, with the point of inflexion at the origin, and is of some physical interest in connexion with the flow produced by a current of air flowing over a region of still air.

6.3. *Definition of the Reynolds Number of the Profiles.*

It is now necessary to define the representative units of velocity and length, and the Reynolds number, appropriate to the profiles.

In each case the representative velocity, U_m, is taken as the maximum velocity of the profile :—

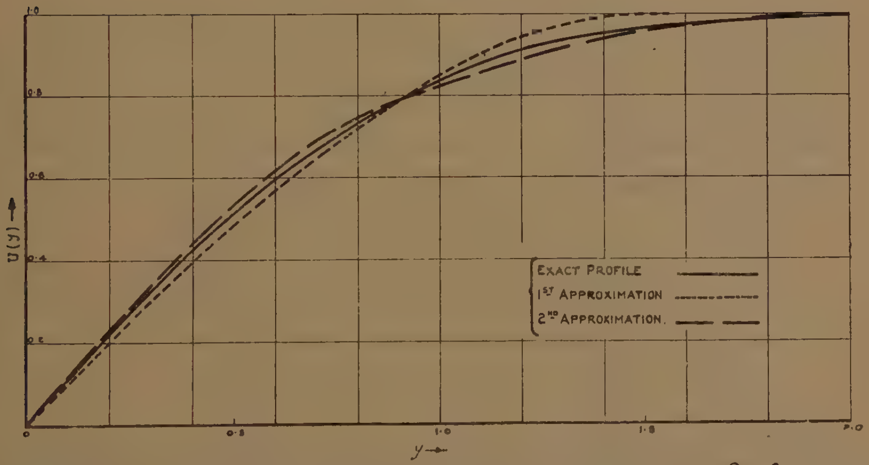
(1) For the profile $U(y)=e^{\frac{1-y^2}{2}}$ we have $U_m=e^{\frac{1}{2}}=1{\cdot}6487$, and we define R_m and h by the equations

$$\left.\begin{aligned}\nu R_m=e^{\frac{1}{2}}h&=\int_0^\infty U(y)\,dy,\\ &=e^{\frac{1}{2}}\sqrt{\frac{\pi}{2}}=2{\cdot}065.\end{aligned}\right\}$$

(2) For the profile $U(y)=\text{erf}\,(y)$ we have $U_m=1$, and we define R_m and h by the equations

$$\begin{aligned}\nu R_m=h&=\int_0^\infty [1-\text{erf}\,(y)]\,dy,\\ &=0{\cdot}567.\end{aligned}$$

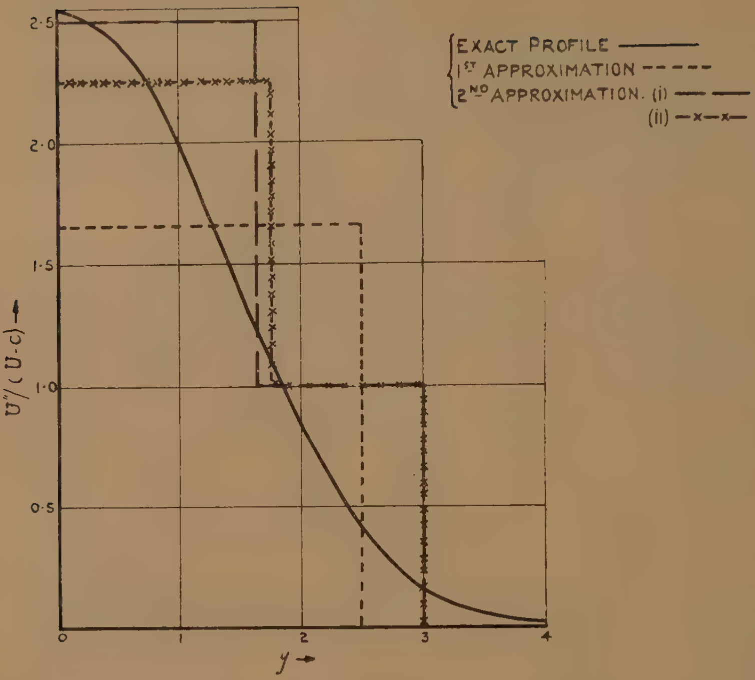
Fig. 2.



Comparison of exact and approximate velocity profiles $U(y)=\frac{2}{\sqrt{\pi}}\int_0^y e^{-y^2}\,dy$.

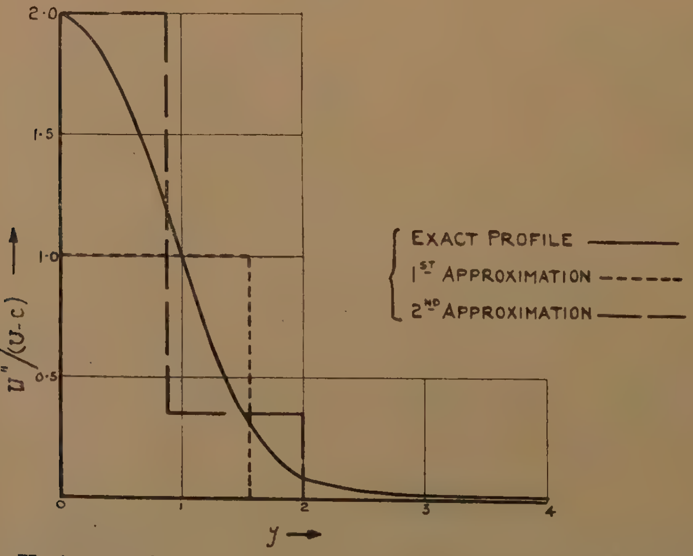
It has already been remarked that the only function of U and its derivatives that occurs in the stability equation (4) is $U''/(U-c)$. Thus it is this function, rather than U or U′, which must be made to approximate as closely as possible to its exact value. Furthermore, the Reynolds numbers of the exact and approximate profiles should be kept as close as possible. Figs. 1–4 show the degree of approximation attained for both profiles, both as regards the shape of the profile itself and also the value of $U''/(U-c)$.

Fig. 3.



Variation of U″/(U—C) with y for exact and approximate profiles $U=e^{\frac{1-y^2}{2}}$.

Fig. 4.



Variation of U″/(U—c) with y for exact and approximate profiles $U=\frac{2}{\sqrt{\pi}}\int_0^y e^{-y^2}dy$.

6.4 (1) *The Velocity Profile* $U(y)=e^{\frac{1-y^2}{2}}$.

Here $U''(y)=0$ when $y=\pm 1$, and so $c=U(\pm 1)=1$.
The boundary conditions for ϕ are

$$\left\{\begin{array}{l}\phi \text{ an even function of } y.\\ \phi\to 0 \text{ as } y\to\infty\ .\end{array}\right.$$

We need only consider the range of positive values of y.

First Approximation.

Take

$$\left.\begin{array}{ll}U=A(1+\cos ky) & (0<y<\pi/k),\\ U=0 & (y>\pi/k),\end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \qquad (19)$$

giving

$$\left.\begin{array}{l}c=U(\pi/2k)=A\\ \\ U''/U-c=-k^2\end{array}\right\}\text{for } 0<y<\pi/k. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (20)$$

and

We have two constants (A and k) at our disposal, and we choose them so that $U(0)$ and $\int_0^\infty U(y)\,dy$ (proportional to Reynolds number) are the same for the exact and approximate profiles.

This gives $\qquad A=0{\cdot}824$

and $\qquad k=1{\cdot}253,$

and the approximate profile is given by

$$\left.\begin{array}{ll}U=0{\cdot}824\,(1+\cos 1{\cdot}253\,y) & (0<y<2{\cdot}505),\\ U=0 & (y>2{\cdot}505).\end{array}\right\}. \quad . \quad . \qquad (21)$$

Equation (4) gives for this profile

$$\left.\begin{array}{ll}\phi''+\phi(1{\cdot}57-\alpha^2)=0 & (0<y<2{\cdot}505),\\ \phi''-\alpha^2\phi \qquad\qquad =0 & (y>2{\cdot}505).\end{array}\right\} \quad . \quad . \quad . \quad . \qquad (22)$$

Solutions of these equations which also satisfy the required boundary conditions are

$$\left.\begin{array}{ll}\phi_1=P\cos\left[\sqrt{1{\cdot}57-\alpha^2}\,.\,y\right] & (0<y<2{\cdot}505),\\ \phi_2=Qe^{-\alpha y} & (y>2{\cdot}505).\end{array}\right\} \quad . \quad . \quad . \qquad (23)$$

Continuity of ϕ and ϕ' at $y=2{\cdot}505$ enable the unknown multipliers P and Q to be eliminated and the following period equation obtained for α :

$$\tan\left[2{\cdot}505\sqrt{1.57-\alpha^2}\right]=\alpha/\sqrt{1.57-\alpha^2}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (24)$$

Solution of this equation by graphical methods gives

$$\alpha=1{\cdot}164. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (25)$$

An alternative form of approximation would be to ignore the condition of equality of Reynolds number and choose A and k so that U and $U''/(U-c)$ are the same at $y=0$ for the exact and approximate profiles.

We then obtain

$$\alpha=1{\cdot}476 \quad \text{and} \quad \nu R_m=1{\cdot}62.$$

Second Approximation.

The second approximation to the exact profile will be of the type :—

$$\left.\begin{aligned} &U=e^{\frac{1}{2}}-A(1-\cos k_1 y) && (0<y<y_0), \\ &U=B[1-\cos k_2 (y_1-y)] && (y_0<y<y_1), \\ &U=0 && (y>y_1), \end{aligned}\right\} \quad . \quad . \quad . \quad (26)$$

where U(0) has been made equal to its value for the exact profile. We have to make U and U′ continuous at $y=y_0$, the continuity condition at $y=y_1$ being automatically satisfied. There are six constants $(A, B, k_1, k_2, y_0, y_1)$ at our disposal, and so some margin of choice is available.

It was considered that when $y>3$, the actual profile might be replaced by the straight line $U=0$. y_1 was therefore taken as 3. The variation of $U''/(U-c)$ with y for the exact profile further suggested taking k_2 equal to unity.

Two choices of the remaining constants were made in order to form some estimate of the sensitiveness of the resulting value of α to small changes in the available constants.

First Choice.

k_1 is so chosen that the point of inflexion occurs at $y=1$, as for the exact profile ; hence $k_1=\pi/2$.

The remaining constants (A, B, and y_0) are determined by the continuity conditions at $y=y_0$, and the condition of equality of Reynolds numbers.

Continuity conditions at $y=y_0$ give

$$\left.\begin{aligned} &e^{\frac{1}{2}}-A\left(1-\cos \tfrac{\pi}{2}\, y_0\right)=B\left(1-\cos \overline{3-y_0}\right), \\ &\tfrac{\pi}{2} A \sin \tfrac{\pi}{2} y_0=B \sin \overline{3-y_0}. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (27)$$

The Reynolds number condition gives

$$\nu R_m=\int_0^{y_0} \left[e^{\frac{1}{2}}-A\left(1-\cos \tfrac{\pi}{2} y\right)\right] dy+\int_{y_0}^{3} B\left(1-\cos \overline{3-y_0}\right) dy\,;$$

hence

$$2{\cdot}065=e^{\frac{1}{2}} y_0-A\left(y_0-\tfrac{2}{\pi} \sin \tfrac{\pi}{2} y_0\right)+B\left(\overline{3-y_0}-\sin \overline{3-y_0}\right). \quad . \quad (28)$$

Graphical solution gives

$$\left.\begin{aligned} y_0&=1{\cdot}65,\\ \mathrm{A}&=0{\cdot}657,\\ \mathrm{B}&=0{\cdot}585,\end{aligned}\right\}$$

and the approximate profile is

$$\left.\begin{aligned} \mathrm{U}&=0{\cdot}992+0{\cdot}657\cos\frac{\pi}{2}y && (0<y<1{\cdot}65),\\ &=0{\cdot}555\,(1-\cos\overline{3-y}) && (1{\cdot}65<y<3),\\ &=0 && (y>3).\end{aligned}\right\} \quad . \quad . \quad . \quad (29)$$

The solutions of equation (4) appropriate to this approximation are

$$\left.\begin{aligned} \phi_1&=\mathrm{P}\cos\sqrt{2{\cdot}467-\alpha^2}\,.\,y && (0<y<1{\cdot}65),\\ \phi_2&=\mathrm{Q}\cosh\,[\sqrt{\alpha^2-1}\,.\,(3-y)]+\mathrm{R}\sinh\,[\sqrt{\alpha^2-1}\,.\,(3-y)] && (1{\cdot}65<y<3),\\ \phi_3&=\mathrm{S}e^{-\alpha y} && (y>3),\end{aligned}\right\}(30)$$

where it is assumed, as indicated by the first approximation, that $1<\alpha^2<2{\cdot}467$.

Furthermore, ϕ and ϕ' must be made continuous at $y=1{\cdot}5$ and $y=3$. Eliminating the ratios of the constants from the resulting equations, we get the period equation for α :—

$$\sqrt{\frac{2{\cdot}467-\alpha^2}{\alpha^2-1}}\,.\tan\,[1{\cdot}65\sqrt{2{\cdot}467-\alpha^2}]=\frac{1+\dfrac{\sqrt{\alpha^2-1}}{\alpha}\tanh\,[1{\cdot}35\sqrt{\alpha^2-1}]}{\tanh\,[1{\cdot}35\sqrt{\alpha^2-1}]+\dfrac{\sqrt{\alpha^2-1}}{\alpha}}, \quad . \quad . \quad . \quad . \quad (31)$$

giving $\alpha=1{\cdot}447$.

Second Choice.

The variation of $\mathrm{U}''/(\mathrm{U}-c)$ with y for the previous approximation (fig. 3) suggested that the value of the constant, k, should be somewhat reduced, and k_1 was accordingly taken as 1·5.

The continuity and Reynolds number conditions give three equations similar to (27) and (28), which yield the values of the constants :

$$\left.\begin{aligned} y_0&=1{\cdot}76,\\ \mathrm{A}&=0{\cdot}689\\ \mathrm{B}&=0{\cdot}520,\end{aligned}\right\}$$

The approximate profile is :—

$$\left.\begin{aligned} \mathrm{U}&=0{\cdot}960+0{\cdot}689\cos 1{\cdot}5y && (0<y<1{\cdot}76),\\ &=0{\cdot}520\,(1-\cos\overline{3-y}) && (1{\cdot}76<y<3),\\ &=0 && (y>3).\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (32)$$

The period equation for α is obtained in a similar manner to that already indicated, thus :—

$$\sqrt{\frac{2{\cdot}25-\alpha^2}{\alpha^2-1}}\tan\left[1{\cdot}76\sqrt{2{\cdot}25-\alpha^2}\right]=\frac{1+\frac{\sqrt{\alpha^2-1}}{\alpha}\tanh\left[1{\cdot}24\sqrt{\alpha^2-1}\right]}{\tanh\left[1{\cdot}24\sqrt{\alpha^2-1}\right]+\frac{\sqrt{\alpha^2-1}}{\alpha}}, \quad (33)$$

giving $\alpha=1{\cdot}378$.

It may be remarked that for this approximate profile the point of inflexion ($U''=0$) is given by $y=1{\cdot}0472$, as compared with $y=1$ for the exact profile.

The above approximations provide some indication of the possible range of values of α; further approximations, being somewhat complex, are not considered, as the above method is to be regarded as primarily of an illustrative nature.

6.5. *The Velocity Profile*

$$U(y)=\frac{2}{\sqrt{\pi}}\int_0^y e^{-y^2}\,dy\equiv \operatorname{erf}(y).$$

We have $U''(y)=0$ when $y=0$, and so $c=U(0)=0$.

The boundary conditions for ϕ are the same as for the previous profile.

First Approximation.

Take

$$\left.\begin{aligned} U&=\sin ky \qquad (0<y<\pi/2k),\\ U&=1 \qquad (y>\pi/2k).\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

The continuity conditions at $y=\pi/2k$ have been automatically satisfied, and we give the remaining constant, k, the value $1{\cdot}007$, so that the Reynolds number is the same as for the exact profile.

Hence the approximate profile is

$$\left.\begin{aligned} U&=\sin 1{\cdot}007y \qquad (0<y<1{\cdot}560),\\ U&=1 \qquad (y>1{\cdot}560),\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (35)$$

and equation (20) gives

$$\left.\begin{aligned} \phi''+\phi(1{\cdot}014-\alpha^2)&=0 \qquad (0<y<1{\cdot}560),\\ \phi''-\alpha^2\phi&=0 \qquad (y>1{\cdot}560).\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (36)$$

Proceeding in a similar manner to that employed for the previous profile, we obtain the period equation for α, thus

$$\tan\left[1{\cdot}560\sqrt{1{\cdot}014-\alpha^2}\right]=\frac{\alpha}{\sqrt{1{\cdot}014-\alpha^2}}, \quad . \quad . \quad . \quad . \quad (37)$$

giving $\alpha=0{\cdot}808$.

If k is taken as $\sqrt{2}$ in equations (34), so that U''/U is the same at $y=0$ for the exact and approximate profiles, we get

$$\alpha=1{\cdot}13 \quad\text{and}\quad \nu R_m=0{\cdot}403.$$

Second Approximation.

There is again some margin of choice for the second approximation, but we shall limit the discussion to profiles of the following type :—

$$\left.\begin{array}{ll} U=P\sin.\sqrt{2}.y & (0<y<y_0), \\ U=\cos k\,(2-y) & (y_0<y<2), \\ U=1 & (y>2). \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

For the exact profile $U=0{\cdot}995$ when $y=2$, and we assume that $U(y)$ may be considered to be unity for larger values of y. U''/U has also been made the same for the exact and approximate profiles at $y=0$.

The three constants (P, k, and y_0) are chosen, as in the "second choice" for the previous profile, to give equality of Reynolds number and to satisfy the continuity conditions at $y=y_0$.

We thus obtain—

$$\left.\begin{array}{l} y_0=0{\cdot}88, \\ k=0{\cdot}60, \\ P=0{\cdot}829. \end{array}\right\}$$

The profile is

$$\left.\begin{array}{ll} U=0{\cdot}829\sin\sqrt{2}k & (0<y<0{\cdot}88), \\ U=\cos\,[0{\cdot}6\,(2-y)] & (0{\cdot}88<y<2), \\ U=1 & (y>2), \end{array}\right\} \quad . \quad . \quad . \quad . \quad (39)$$

and the resulting period equation for α is

$$\sqrt{\frac{2-\alpha^2}{\alpha^2-0{\cdot}36}}.\tan\left[0{\cdot}88\sqrt{2-\alpha^2}\right]=\frac{1+\dfrac{\sqrt{\alpha^2-0{\cdot}36}}{\alpha}\tanh\left(1{\cdot}12\sqrt{\alpha^2-0{\cdot}36}\right)}{\tanh\left(1{\cdot}12\sqrt{\alpha^2-0{\cdot}36}\right)+\dfrac{\sqrt{\alpha^2-0{\cdot}36}}{\alpha}} \quad . \quad . \quad . \quad . \quad (40)$$

giving $\alpha=1{\cdot}09$.

Various other choices for the second approximation were considered, but the resulting changes in the value of α were small, and it was not thought necessary to give further details.

7. *Symmetrical Oscillations.*

We can now return to the assumption that symmetrical oscillations do not produce instability.

It can easily be shown that this assumption is correct for the simple case of the first approximation to either of the above profiles, and the method of proof suggests that this conclusion will apply to profiles in general.

The approximation can be considered as one of the following types :

$$\left.\begin{aligned} &U=A(1+\cos ky) && (0<y<\pi/k), \\ &U=0 && (y>\pi/k), \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (41)$$

for the symmetrical profile ; or

$$\left.\begin{aligned} &U=\sin ky && (0<y<\pi/2k), \\ &U=1 && (y>\pi/2k), \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

for the " monotonic " profile.

In either case, the appropriate *odd* solution for ϕ is

$$\left.\begin{aligned} &\phi_1=P\sin\sqrt{k^2-\alpha^2}\,.\,y && (0<y<y_1), \\ &\phi_2=Qe^{-\alpha y} && (y>y_1), \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (43)$$

where y_1 equals π/k in the first case and $\pi/2k$ in the second.

The period equation for α, assuming $\alpha<k$, is

$$\tan\left[\sqrt{k^2-\alpha^2}\,.\,y_1\right]=-\sqrt{k^2-\alpha^2}/\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

Now since α and y_1 are both positive, the left-hand side of this equation will always be positive and the right-hand side will always be negative. (The degenerate case of $\alpha=k$ is excluded, since this would lead to the evanescence of ϕ_1.)

There is thus no real solution for α, and hence instability cannot be produced by oscillations of this type.

A similar argument, involving hyperbolic functions, can be applied when $\alpha>k$.

8. *Summary*.

The results obtained for the two profiles considered may be summarized in Table I.

TABLE I.

Summary of Oscillation Calculations for Laminar Profiles.

Profile.	$U(y)=e^{\frac{1-y^2}{2}}$.	$U(y)=\frac{2}{\sqrt{\pi}}\int_0^y e^{-y^2}\,dy$.
Wave-velocity, c	1	0
Reynolds number, R_m	2·065	0·567
Approximate value of α :—		
First approx. (1)	1·164	0·808
(2)	1·476 (R_m=1·62)	1·13 (R_m=403)
Second approx. (1)	1·447	1·09
(2)	1·378	—

The treatment of the two particular profiles should be regarded primarily as an illustration of the type of result to be obtained from the basic hypothesis expressed in equation (16). The methods of approximation can clearly be refined and extended, but it has not been considered necessary to do so, as the nature of the methods is sufficiently indicated by the results given.

Their further interpretation will be considered in Part II. of this paper, in which a comparison with experiment is attempted.

PART II.—EXPERIMENTAL INVESTIGATION OF THE WAKE AND COMPARISON WITH THEORY.

9. *Introduction.*

By the "wake" behind a solid body moving through a fluid is understood the region in which the "smooth" flow of the fluid is disturbed. Many workers have directed their attention to this part of the field because of its bearing on the problem of fluid resistance; in an inviscid fluid there would be no wake, and the moving body would experience no resistance to its motion. To account for the state of affairs actually observed, frictional shearing forces must be taken into consideration, and the conception of the "boundary layer" provides the usual method by which this is achieved. The types of flow that are observed in the wake are of great variety and complexity, and no single theory can be expected to provide a valid approximation over more than a limited range of values of the Reynolds number.

A rough outline of the sequence of changes that occur in the rear of a streamline body when the Reynolds number, R, is progressively increased is as follows :—

(1) For very small values of the Reynolds number the closing-up of the streamlines behind the body is very nearly complete. This type of motion has been termed "creeping motion" *, and the approximations of Stokes and Oseen may be employed in the theoretical treatment.

(2) There follows a range of R in which the motion is steady and the wake broadens slowly downstream. In this range the approximations of the boundary-layer theory provide a close approach to the observed flow.

(3) As R is increased the wake begins to oscillate, assuming first a sinusoidal form. Discrete vortices, arranged more or less in the asymmetrical "Karman Street" formation, tend to form in the hollows of the

---

* Prandtl-Tietjens, 'Fundamentals of Hydro- and Aeromechanics.' English translation by L. Rosenhead, p. 261. (McGraw-Hill Book Co., 1934.)

undulations. The proportion of the total vorticity of the fluid which is concentrated in these vortices increases with R, but a residuum of vorticity always remains in the sinusoidal wake between the vortices.

(4) On further increasing R discrete vortices are shed immediately behind the moving body of the fluid.

A periodic fluctuation of the vorticity and velocity in a wake of this nature can be detected over a considerable range *.

(5) At Reynolds numbers of the order of 10^6 a state of " fully developed turbulence " is attained.

It is not possible to give numerical values for the Reynolds numbers appropriate to these various types of flow, on account of the great influence of the shape of the moving body. The general statement can, however, be made that the " smoother " (or more " streamline ") is the body, the larger is the range of Reynolds number which gives rise to any one of the above types of wake.

There is some divergence among writers on the subject as to the use of the word " turbulent " to characterise fluid flow. It is here suggested that the word be used only in the sense of the " fully developed turbulence" of the range (5) above. Here all trace of periodic eddies is lost : small random oscillations of velocity are superposed on the main flow, and experiments by Fage † with the ultramicroscope have shown that these fluctuations are three-dimensional. Many writers apply the word " turbulent " to the régime (4) and even (3) above, whereas " periodic " or " eddying " would appear to be a more appropriate designation.

10. *Apparatus and Experimental Method.*

The experiments were carried out in the research water tank at the Aeronautics Department, Imperial College of Science and Technology, South Kensington, the main object being to obtain photographs of the velocity field in the wake behind a body moving through still water. By analysis of these photographs quantitative information can be obtained as to the characteristics of the wake. A cylindrical model, whose generating lines are horizontal, is drawn through still water in a horizontal direction at right angles to the generating lines of the cylinder. The central vertical plane of the tank is illuminated from above, and photographs of the movement of particles suspended in the water are taken by a camera at the side. The duration of the exposure is made sufficiently short, so that the traces of the particles on the photographic plate may be assumed to define the instantaneous velocity vectors of the fluid field.

---

* Fage and Johansen, Proc. Roy. Soc. A, cxvi. p. 170 (1927).

† Aero. Res. Comm. Reports and Memo. no. 1520 (1932).

A general description of the apparatus and an account of the experimental technique has been given by Richards *, a previous worker on the tank, and so need not be repeated. Various modifications have been made in the apparatus, however, of which the following are worthy of mention.

The carriage, to which is attached the model, now runs on rails above the tank. The valve at the back of the tank, through which passes the spindle, was removed. There is a secondary tank behind the main tank, also filled with water, and divided from the main tank by a partition in which there is a $\frac{3}{4}$-in. slot along the whole length of the tank. A spindle projects from the back of the model, through this slot into the rear tank, and is connected to the carriage. A large shield is fixed to the spindle, so that the slot is covered in the neighbourhood of the model. There is thus no effective disturbance of the water in the main tank due to the travel of the spindle and carriage connexions along the secondary tank.

The hydraulic driving mechanism has been replaced by a 1 h.p. electric motor which turns a " lead screw " (worm-drive) of $\frac{1}{6}$-in. pitch, running the whole length of the tank above the water. The carriage is driven by means of a nut engaging on the lead screw. A potentiometer device enables the speed of the model through the water to be controlled. Considerable difficulty was experienced in obtaining a sufficiently constant speed during the whole run along the length of the tank. Very small departures from the geometrical spiral of the lead screw produced a marked variation in speed. The nut was reduced in size and allowed " play " in an axial direction, and the final experiments were made when the speed variation did not exceed 4 per cent. This is far from satisfactory, and it is to be hoped that further improvement will be possible in this direction.

The chief advance was made in the method of recording the velocity of the model and the instant and duration of the exposure. A " fractional seconds chronograph " was used, a very delicate instrument invented by Mr. E. A. Nehan, which can record four sets of " events " in addition to the time. Automatic contacts actuated by the moving carriage were fitted so that four photographs could be taken on a single run, thus showing the development of the wake up to a distance of about 4 feet behind the model. In the later experiments the contacts were spaced so that a continuous picture of the wake was obtained when the

---

* Phil. Trans. A, ccxxiii. p. 279 (1934).

camera was placed so as to give unit magnification. The photographic plates were of full size, so that a distance of about 30 inches behind the model could be photographed.

The lighting and photographic arrangements have been retained in their original form, although experiments were made in the use of two "line-filament" G.E.C. lamps.

The contacts were $\frac{1}{2}$ in. wide, hence the duration of the exposure is the time taken for the model to travel this distance. The revolving shutter was eliminated, so that the traces of the photographic plate were unbroken during the exposure. The platinum wire used by Richards was also dispensed with, as the position of the model for a particular photograph could be determined immediately from the position of the appropriate contact. The velocity of the model was determined by a spring contact so arranged as to give a marking on the chronograph record at every revolution of the lead screw.

The substance used to render the flow visible was a mixture of unsweetened condensed milk and alcohol. Considerable trouble, as noted by Richards, was experienced in obtaining a field of milk-particles which would remain suspended in the water for the three or four hours necessary to eliminate all random motion. Careful photographic technique was required to produce a plate on which the milk-traces were sufficiently clear and sharp for accurate measurement with the travelling microscope. The dual problem of the particle suspension and photographic technique still presents the chief obstacle to the rapid and certain use of the apparatus. The slow model speeds make it essential to wait for some hours after the introduction of the milk particles before taking photographs, and a very sensitive density balance must be achieved.

11. *Experiments Performed and General Remarks.*

Two models were used to produce the wake, a flat plate and a symmetrical aerofoil section, both at zero incidence. The flat plate was of $1\frac{1}{2}$ in. chord, $\frac{1}{16}$ in. thick at its centre, and tapered smoothly to sharp edges. There was one small speed range within which resonance effects produced considerable vertical vibration of the plate. These speeds were, of course, avoided, and otherwise no appreciable vibration occurred. The aerofoil section of 4-in. chord was that known as "R.A.F. 30." The wake is much wider than that produced by the flat plate, but is essentially similar in configuration and development.

The speed range was determined by the apparatus: appreciable variations of the model speed occurred during the run, and the residual motion of the water became comparable with that produced by the model for speeds less than $\frac{1}{2}$ cm./sec., while above 4 cm./sec. the photographic plates were insufficiently clear and general vibration tended to occur.

Let $R = \frac{U_0 L}{\nu}$ denote the Reynolds number of the model, where U_0 is the velocity and L is the chord of the model, and ν is the kinematic viscosity of the water. Then the available speed range corresponds to a range of R of about 170–1400 for the flat plate and 450–3600 for the aerofoil.

Following Richards' investigation * it is assumed that the motion is entirely two-dimensional.

The types of wake which are produced in these ranges of R fall in the divisions (2) and (3) of the Introduction (§ 9). It is therefore possible to study the first departure of the wake from the laminar form, and the early development of the unstable sinusoidal oscillations which at higher Reynolds numbers break up to form the discrete vortices of the " Karman Street " type. The onset of instability is known to be very sensitive to small residual turbulence in the fluid, and great care was taken, as has already been stated, to allow the water to come to rest before beginning an experiment. The method here adopted of moving the disturbing body through stationary fluid already has great advantages in this respect over the more usual method of placing a fixed body in a current of fluid. In a recent study by Rosenbrook † of the instability of the friction layer in a slightly diverging channel the experimental arrangements were of this latter type, and it is possible that his results are vitiated to some extent by the turbulence that must inevitably have been present in the channel.

Some general results revealed by the experiments may first be noted. The laminar wake does not show any signs of oscillation throughout the length of the tank until a Reynolds number is reached of about 600 in the case of the flat plate and about 1000 in the case of the aerofoil section. The widening of the wake downstream is very small, due in part to the compressive action of the walls of the tank. There is then a range in which the commencement of the oscillation depends very largely on the very small residual motion of the water, and no definite limit can be given for the transition. Frequently the oscillation persists for a few wave-lengths, then dies out, and the wake becomes sensibly laminar again. The effect may recur a short distance further downstream. The wave-length of these " pseudo-oscillations " is remarkably constant, but the amplitude varies considerably and appears to depend almost entirely on the small residual motions.

The slowest speed for which photographs of the wake behind a flat plate were obtained corresponded to a Reynolds number of 409. In this case the wake was entirely laminar, while for $R=636$ a small trace

---

* *Loc. cit.*

† G. Rosenbrook, *Zeit. für Angew., Math. u. Mech.* xvii. no. 1 (Feb. 1937)

of oscillation was detected at some distance behind the plate. Pl. IX. shows this latter wake immediately behind the plate. The wake for R=791 furnished a good example of the "pseudo-oscillations." Two sets of photographs corresponding to R=1090 and 1110 showed well-defined oscillating wakes, although the amplitude of the oscillations was observed to vary irregularly in both cases, indicating the effects of small residual motions. Pl. X. shows the wake for R=1090 at a distance of 25 cm. behind the plate. The "pseudo-oscillations" effect was much less marked in the wakes produced by the aerofoil section.

At higher R the oscillations do not show any tendency to die away, and the wake shows a regular sinusoidal form. The wave-length remains constant, and the amplitude may increase slightly or may remain sensibly constant. The development of such oscillations is shown in Pls. XI.–XIII., which form a sequence of photographs of the wake behind the aerofoil section at a Reynolds number of 1370, up to a distance of about 55 cm. behind the model.

At large distances downstream the effects of diffusion begin to be felt and the amplitude decreases slightly, but the wave-length remains unaltered. The tendency for the sinusoidal wake to break up into isolated vortices does not become manifest until a larger Reynolds number is reached, which may be placed at about 1300 for the flat plate and 1800 for the aerofoil. It is usually possible to obtain a "Karman Street" for a distance downstream of at least 10L before diffusion destroys the regularity.

At higher values of R, although instability is greater, diffusion has a considerable effect and the vorticity disappears fairly rapidly, and a variety of results are obtained. Sometimes a few vortices are visible, at other times there is diffusion almost immediately behind the model. It may be estimated that a Reynolds number of about 2000 for the flat plate and 2500 for the aerofoil section represents the upper limit above which the vortex street cannot be said to exist. Unfortunately the speeds required to give these Reynolds numbers are too high to enable clear photographs to be obtained, and the conclusions have to be drawn from visual observations.

These general observations have been elaborated in some detail because the opinion appears to be prevalent that both the commencement of instability and the rapid diffusion of the "Karman vortices" occur at lower Reynolds numbers than have here been observed. The delayed instability here obtained may be due to the more complete exclusion of residual turbulence.

The quantitative results obtained from analysis of the photographs will now be considered. Only the clearest photographs were selected for detailed measurement, while the general remarks above are based on inspection of the aggregate of the photographs obtained. Attention will now be confined to the plate-measurements.

Table II. gives some particulars of the plates that were measured in order to obtain the velocity distribution in the wake.

TABLE II.

Particulars of Plates measured to give Velocity Distribution in the Wake.

No. of plates.	Ref. nos. of plates.	U_0 (cm./sec.).	ν (cm.^{2}—sec.).	R.	Type of wake.
		Flat plate. L=3·81 cm.			
4	F. 13–16	1·226	0·0115	409	Laminar.
3	F. 5–7	1·905	0·01095	636	Very slight "pseudo-oscillations."
		Aerofoil section. L=10·16 cm.			
4	—	0·658	0·0106	631	Laminar.
4	—	1·485	0·0110	1370	Sinusoidal.

In §18 particulars will be given of other plates that were measured to obtain the characteristics of oscillating wakes.

12. *Notation.*

The theoretical investigations which are most suitable for comparison with the experimental results are those of Goldstein *, and in order to facilitate comparison the notation of his paper (G. II.) will be adopted as far as possible, rather than that of Part I of this paper.

The following is the notation adopted :—

U_0, L, ν, R. Already defined.

X, Y=Cartesian coordinates with the X-axis downstream. The origin is on the line of symmetry, but its exact position is unknown.

W, V=velocity components in the wake.

$U=U_0-W$.

$u=U/U_0$.

---

* Proc. Camb. Phil. Soc. xxvi. p. 18 (1930), quoted subsequently as G. I.; Proc. Roy. Soc. A, cxlii. p. 545 (1933), quoted subsequently as G. II. ; *ibid.* cxlii. p. 563, quoted subsequently as G. III.

$w = \mathrm{W}/\mathrm{U}_0$.
$x = \mathrm{X}/4\mathrm{L}$.
$y = \mathrm{Y}/\mathrm{L}$.
$\eta = \left(\frac{\mathrm{U}_0}{2\nu\mathrm{X}}\right)^{\frac{1}{2}}\mathrm{Y}$.
x_1 = distance downstream from rear edge of body.
c = a constant equal to $(\mathrm{X}-x_1)/\mathrm{L}$.
D = drag per unit breadth of the body.
k_{D} = a drag coefficient equal to $\mathrm{D}/\rho\mathrm{U}_0{}^2\mathrm{L}$.

The symbols above which are different from those used by Goldstein are k_{D} and y.

13. *Comparison with Theory.*

A theoretical treatment of the velocity distribution at a considerable distance behind a moving body (the motion being assumed steady) is given by Goldstein in G. II. He uses the conception of the boundary layer and also the Oseen approximation : w is assumed small when x is large, and it is further assumed that w can be expanded in the form

$$\frac{f_1(\eta)}{x^{\frac{1}{2}}} + \frac{f_2(\eta)}{x} + \frac{f_3(\eta)}{x^{\frac{3}{2}}} + \ldots \qquad (45)$$

Only two approximations can be derived giving f_1 and f_2, a third approximation not being obtainable by these methods. The results obtained are :—

$$f_1(\eta) = \frac{\beta}{\sqrt{\pi}} e^{-\frac{1}{2}\eta^2}, \qquad (46)$$

$$f_2(\eta) = \frac{\beta^2}{2\pi}\left[e^{-\eta^2} + (\tfrac{1}{2}\pi)^{\frac{1}{2}} . \eta e^{-\frac{1}{2}\eta^2} \operatorname{erf}(\eta/2)\right], \qquad (47)$$

where β is a numerical constant, given by momentum considerations as

$$\beta = \tfrac{1}{4}k_{\mathrm{D}}\mathrm{R}^{\frac{1}{2}}. \qquad (48)$$

It should be remarked that the origin of x is at present unknown, that these results are only valid at a considerable distance downstream, and that no restrictions have yet been placed on the type of body producing the wake.

It might be thought that the various assumptions on which the calculations are based are mutually contradictory. Goldstein gives a discussion of this matter, and suggests that the various features of the transition to turbulence—or, as we should prefer to style it, the transition to unsteadiness in the wake (régime 3 of the Introduction)—should be the subject of experimental investigation. The results given in this paper provide a partial answer to these questions. It has already been stated that the laminar wake persists up to a larger Reynolds number than is generally recognized, and the photographs show wakes which remain laminar downstream as far as can be observed, and for which the Reynolds number is large enough for the conception of the boundary layer to be applied. The validity of the third assumption, whether w is sufficiently small for the method of approximation to be valid, can only be tested by a comparison of the calculated and experimental results.

For the particular case of the flat plate, the numerical evaluation can be carried out by making use of the drag formula obtained by Blasius * on the basis of the boundary-layer conceptions. The value of the constant β is thus found to be 0·33204. In a previous paper † the velocity in the wake just behind the rear edge of a flat plate (x_1/L less than 0·5) was calculated by means of series solutions of the boundary layer equations, on the assumption of zero pressure gradient downstream. Thus the velocity has been calculated throughout the whole wake, and it is found that the velocity curves join up smoothly if the origin from which X is measured is taken at a distance of 0·52L upstream from the rear end of the plate—in our notation, $c=0{\cdot}52$.

It is with the comparison of these calculated velocity distributions with experiment that we shall now be concerned.

The only experimental results so far published in which a comparison is made are those of Fage ‡. Comparison is made just behind the rear edge of the flat plate ($x_1/\mathrm{L}=0{\cdot}02$ and $x/\mathrm{L}=0{\cdot}512$) for Reynolds numbers of $0{\cdot}54\times10^5$ and $2{\cdot}01\times10^5$ in each case. The results indicate that at these high values of R the wake, whilst still laminar for $x_1/\mathrm{L}=0{\cdot}02$, has become unsteady when $x_1/\mathrm{L}=0{\cdot}512$. It will be seen at once that these experiments only provide a comparison with the results of G. I., and provide no indication of the validity of the methods employed in G. II. The present photographs do not permit of accurate measurement just behind the rear edge of the plate, and the comparison is made for values of x_1/L ranging from about 1 to 20, *i. e.*, in the range covered by the theory of G. II.

14. *The Velocity Field in the Laminar Wake behind a Flat Plate.*

Two sets of photographs of the wake behind a flat plate moving tangentially were measured corresponding to Reynolds numbers of 409 and 636. General data for these two cases are given in Table II. The velocity distribution was obtained along the line of symmetry ($\mathrm{Y}=0$) of the wake, and also along various lines at right angles ($\mathrm{X}=$constant). In all cases the velocity in the Y-direction was very small, and no attempt was made to measure it.

---

* *Z. Math. Phys.* lvi. p. 1 (1908).

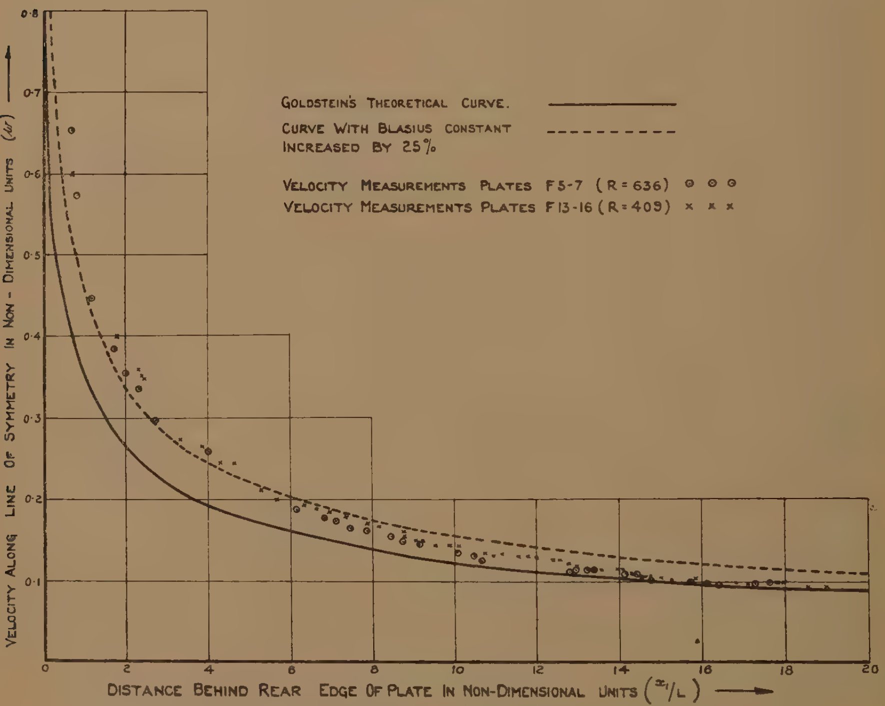
† G. I.

‡ Appendix to G. II.

14.1. *Velocity along Line of Symmetry.*

The X- and Y-coordinates of the mid-points of the particle tracks were measured, and the magnitude of the velocity at that point in the field was given by the length of the track, obtained by means of a travelling microscope. For details of the method of analysis of the photographic plates reference may be made to Richards *. A tolerance of ±0·2 cm. was allowed on each side of the line of symmetry.

Fig. 5.



Velocity along the line of symmetry of the wake. Flat plate.

It will be seen from equations (45) to (47) that Goldstein's calculated velocity w along the line of symmetry is independent of the Reynolds number of the plate.

Fig. 5 shows the calculated velocity distribution down the centre line, and also the observed values for the two cases R=409 and 636. It will be seen that the experimental points lie above the theoretical curve, and that the deviation is somewhat greater for the lower Reynolds number.

---

* *Loc. cit.*

Goldstein's results are based, as already stated, on the Blasius drag formula, which is of an asymptotic nature for large values of R. It is generally known that the value of $k_D R^{\frac{1}{2}}$ (hereafter denoted by K) increases as R is reduced. Fage * has carried out some experiments to obtain the variation of this quantity with R. The results are still somewhat scanty, but present knowledge indicates that for $R=409$ (the value for plates F. 13–16) the value of K (or β) should be increased by about 25 per cent., and for $R=636$ (plates F. 5–7) very slightly less. Fig. 5 therefore also shows the velocity distribution obtained by Goldstein's method when this increase has been effected. Alteration of the value of β will increase the value of c required for the two theoretical curves to join up smoothly, and hence cause a displacement of the whole curve. The change will be small, however, and it has not been thought necessary to revise the whole computation to take account of this effect. The matter will be considered again in connexion with the aerofoil section results.

The agreement between the observed and calculated velocities may be said to be improved by the increase in the value of β, and is closest for the middle range of values of x_1/L. For small values of x_1/L the method of G. II. is not applicable, and the effect of the displacement of the theoretical curve will be appreciable. For large values of x_1/L the observed velocities fall below the calculated curve, and tend to approximate to the original Goldstein curve. In general, it may be stated that the slope of the experimental curve is (algebraically) greater than that of the calculated curve, this being probably due to the influence of the walls of the tank.

In paper G. III. Goldstein compares his results with those of Filon †, who calculated the velocity distribution behind a solid body without making the boundary layer assumptions. Goldstein has shown that his results differ from Filon's only by the omission of an extra term in the velocity formulæ. For $\eta=0$ a term of value $\frac{1}{x}\frac{\beta}{2\pi\sqrt{R}}$ has to be subtracted from the value of w given in G. II. For the two cases considered, the magnitude of this term is of the order of $0{\cdot}003/x$, which may clearly be neglected. A partial justification of the validity of the boundary-layer assumptions, for the range of R considered, is thus provided.

14.2. *Velocity Distribution along Lines at Right Angles to the Line of Symmetry.*

Measurements were also made of the velocity distribution along lines $X=$constant for the two wakes.

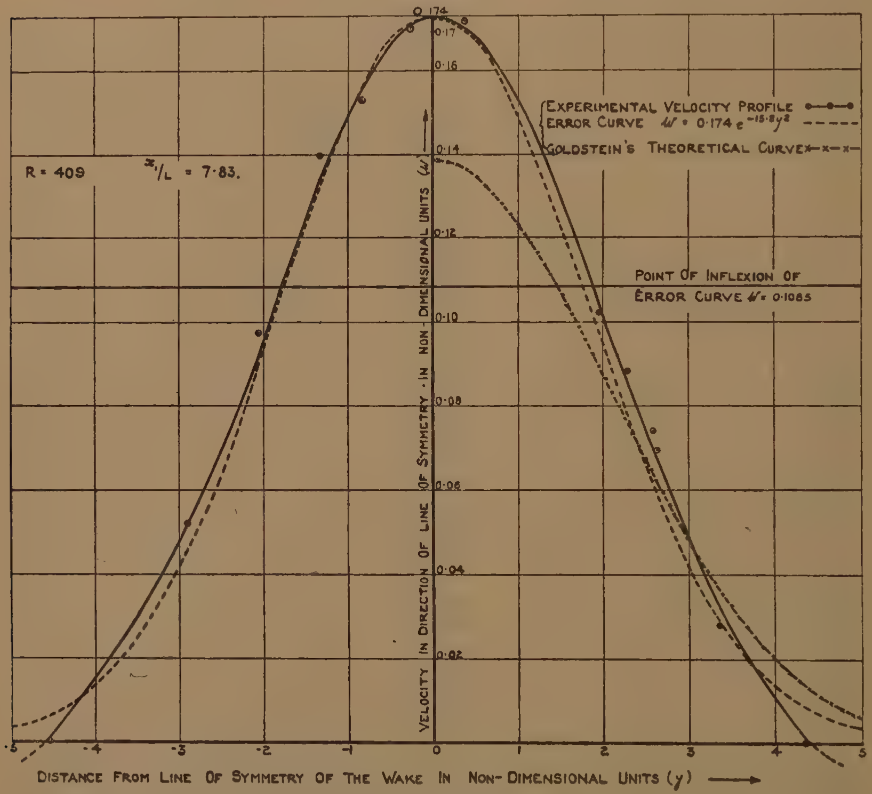
---

* Aero. Res. Comm. Reports and Memo. no. 1580 (1933).

† Proc. Roy. Soc. A, cxiii. p. 7 (1926); Phil. Trans. A, ccxxvii. p. 93 (1928).

Figs. 6 and 7 show specimen velocity profiles for the wake at distances downstream given by $x_1/L=7{\cdot}83$ and $12{\cdot}0$, at a Reynolds number of 409 (plates F. 13–16), together with the theoretical profiles obtained in G. II., and corresponding to the Blasius value of β. It will be seen that the agreement would be improved if β were increased by about 25 per cent., as discussed above, but the requisite computations have not been made as the general nature of the comparison is apparent from the figures.

Fig. 6.



Velocity at right angles to the line of symmetry of the wake. Flat plate.

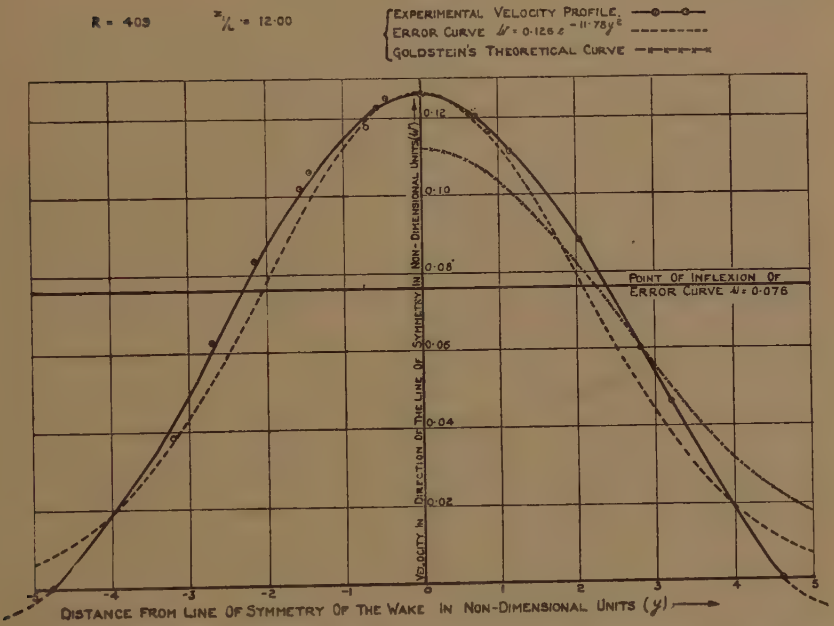
The chief differences between the two sets of curves occur at the edges of the wake. The theoretical calculations apply to an unlimited field of fluid, while the experiments were made in a tank of finite size. The photographs indicate sharp " edges " to the wake, given by two lines of zero velocity. Outside these limits there is a small " backflow," too small to be accurately measured, and it is seen at once, from continuity considerations, that the value of $\int w\,dy$ across the whole depth of the tank must be zero. The general shape of the experimental velocity curves also exhibits the compressive action of the walls of the tank.

15. *Experimental Calculation of the "Drag" of the Flat Plate.*

Considerations of momentum show that in an unlimited fluid the drag coefficient is given by

$$k_D=\int_{-\infty}^{\infty} w(1-w)\,dy \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

Fig. 7.



Velocity at right angles to the line of symmetry of the wake. Flat plate.

(integral along a line x_1=constant, behind the body), where the pressure-gradient downstream is taken as zero to a first approximation. (For a discussion of this assumption see G. III.)

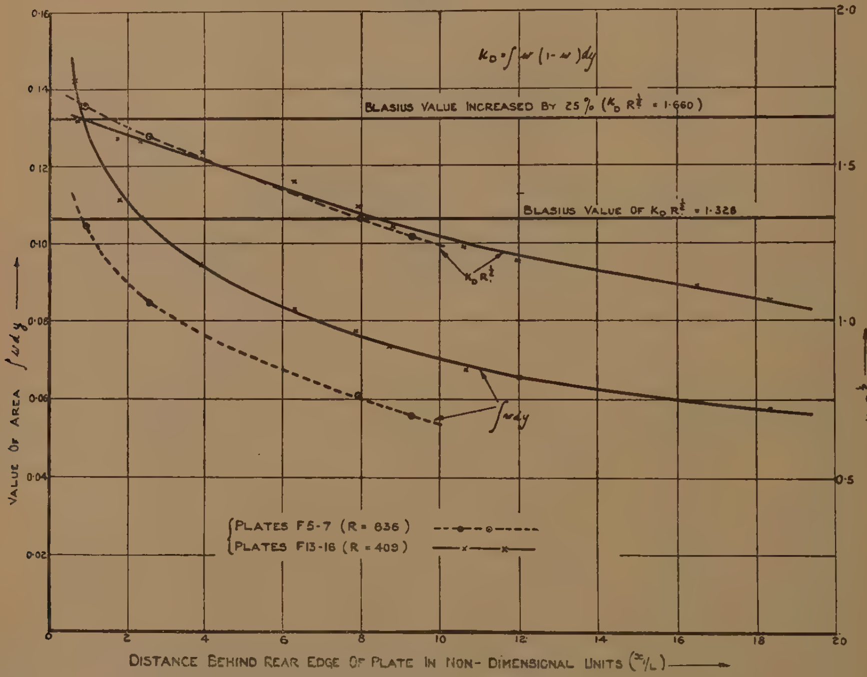
If the integral is taken at an infinite distance downstream, $\int_{-\infty}^{\infty} w^2\,dy \to 0$, while $\int_{-\infty}^{\infty} w\,dy$ remains finite, and the above equation reduces to

$$k_D=\int_{-\infty}^{\infty} w\,.\,dy_{(x\to\infty)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

This is the result used in G. II. to evaluate the numerical constant, β, for the case of a flat plate.

These equations require modification in the case of a limited body of fluid bounded by parallel walls. A fictitious "drag coefficient" may be defined by equation (49), where the integrals are evaluated across the wake (*i. e.*, between the points where the velocity is zero), and the backflow is neglected. The values of $\int w\,dy$ and $\int w^2\,dy$ were calculated by graphical integration for the measured velocity profiles, and the results (together with the fictitious value of k_D) shown in Table V. Fig. 8 shows the variation of $\int w\,dy$ with x_1/L for the two sets of plates. In each case the value of $\int w\,dy$ decreases downstream and the curve for R=636 lies wholly above that for R=409.

Fig. 8.



Variation of $\int w\,dy$ and $k_D R^{\frac{1}{2}}$ with x_1/L. Flat plate.

It is a general feature of all theoretical results that $K(=k_D R^{\frac{1}{2}})$ is independent of both R and x_1/L. For example,

Bairstow, Cave, and Lang * obtain K=3·054 for R=4.
Southwell and Squire † ,, K=3·126 for R=4.
Piercy and Winny ‡ ,, K=2·257 as an asymptotic result when R is large.

---

* Phil. Trans. A, ccxxiii. (1923).

† Phil. Trans. A, ccxxxii. (1933).

‡ Proc. Roy. Soc. A, cxl. (1933).

All these workers used Oseen's approximation or some modification of it.

The calculated values of K are plotted on fig. 8. The curves for the two values of the Reynolds number are seen to " join up " fairly well, thus indicating that K is independent of the Reynolds number. There is, however, a marked decrease of K with x_1/L, which may be mainly attributed to the effect of the walls of the tank. The value of 1·328 for K obtained by Blasius (used in G. II. and denoted by 4β) is shown in fig. 8, and also the value obtained by increasing β by 25 per cent., as already discussed. The experimental value of K is in close agreement with this augmented value immediately behind the plate.

TABLE III.

Summary of Area and Drag Calculations for Laminar Flat Plate Wakes.

x_1/L.	$\int w\,dy$.	$\int w^2\,dy$.	$\int w^2\,dy$ as % of $\int w\,dy$.	$k_{\mathrm{D}}\mathrm{R}^{\frac{1}{2}}$.
Plates F. 13–16. R=409.				
0·695	0·1420	0·0600	42·2	1·659
1·746	·1116	·0330	29·6	1·591
2·34	·1067	·0281	26·3	1·590
3·90	·0945	·0182	19·3	1·540
6·29	·0836	·0124	14·8	1·440
7·83	·0776	·0101	13·8	1·365
8·75	·0735	·0086	11·7	1·308
10·64	·0677	·0068	10·05	1·232
12·00	·0653	·0065	9·95	1·190
16·60	·0598	·0045	7·52	1·118
18·26	·0572	·0039	6·75	1·072
Plates F. 5–7. R=636.				
0·973	·1044	·0381	37·0	1·692
2·59	·0849	·0226	26·6	1·591
7·93	·0615	·0094	15·3	1·330
9·27	·0560	·0062	11·07	1·270

The variation of K with x_1/L serves, in a general qualitative manner, to explain the results exhibited in fig. 5. The value of β appropriate to the observed results is approximately 25 per cent. above the Blasius value just behind the plate, but decreases downstream until its value at a considerable distance from the plate is less than the Blasius value. The velocities along the line of symmetry of the wake should therefore begin by approximating to those given by the upper curve of fig. 5, fall gradually below this curve until the lower curve is approached, and eventually crossed. The experimental points of fig. 5 show that this is, in fact, the case.

The first approximation to the velocity profile for any section $x_1=$ constant, given in G. II., is the Gaussian error curve.

The velocity may be written approximately

$$w=w_m e^{-py^2}, \quad \ldots \ldots \ldots \ldots \quad (51)$$

where w_m is the maximum velocity of the section. Both w_m and p will vary with x_1; while p will also be proportional to the Reynolds number of the flow if it be assumed that $R^{\frac{1}{2}} \int_{-\infty}^{\infty} w\,dy$ is independent of R.

In connexion with the comparison between experiment and the theory of Part I., it is of importance to consider the validity of taking a velocity profile of the Gaussian form as an approximation to the velocity distribution along any section $x_1=$constant in the actual wake; and in figs. 6 and 7 Gaussian error curves are drawn having the same maximum ordinates, w_m, and the same areas as the experimental velocity profiles. ($\int_{-\infty}^{\infty} w\,dy$ for the Gaussian profile is equated to $\int w\,dy$ across the wake for the experimental profile.) The agreement may be considered satisfactory, the greatest divergence occurring at the edges of the wake, due to the "backflow" already discussed.

16. *Velocity Field in the Wake behind an Aerofoil Section.*

Similar measurements were made on the wakes formed behind the symmetrical aerofoil section previously described. Two sets of photographic plates were analysed; one, corresponding to a Reynolds number of 631, showed a purely laminar wake, while the other (R=1370) exhibited a sinusoidal wake. Further data have been given in Table II.; Pls. XI.–XIII. show a sequence of three photographs of the oscillating wake.

16.1. *Laminar Wake.* R=631.

The measurements made for the laminar wake are exactly similar to those already described in the case of a flat plate.

We have velocity of model, $U_0=0{\cdot}657$ cm./sec., and chord of model, $L=10{\cdot}16$ cm.

The results for the velocity along the line of symmetry are plotted in fig. 9. The values of $\int w\,dy$ and $k_D R^{\frac{1}{2}}$ are obtained from the experimental measurements in the manner already described, and the results are plotted in fig. 10 and tabulated in Table IV.

TABLE IV.

Summary of Area and Drag Calculations for Laminar Aerofoil Wakes.

Plate.	x_1/L.	w_m.	$w\,dy$.	K=$k_D R^{\frac{1}{2}}$.
1	0·423	0·74	0·1343	1·87
2........	1·24	·51	·1112	1·81
	1·57	·43	·1004	1·76
	1·82	·40	·0915	1·65
	2·11	·36	·0817	1·59
	2·77	·295	·0708	1·40
3........	3·34	·27	·0654	1·33
	3·80	·24	·0602	1·248
	4·25	·225	·0566	1·20
	5·00	·195	·0492	1·07
4........	6·51	·170	·0443	0·99
	6·23	·155	·0403	0·902
	6·83	·140	.0382	0·865

Goldstein's theoretical results, as applied to the wake behind an aerofoil section, are still expressed by equations (45) to (48). The Blasius result is no longer available to determine the numerical value of the constant, β, which must now be put equal to $\frac{1}{4}.k_D R^{\frac{1}{2}}$. The calculations given in G. I. for the field just behind a flat plate are not applicable, and the value of c cannot be taken as 0·52, but is theoretically indeterminate. The values of c and β must accordingly be obtained from the experimental results.

For the line of symmetry, $\eta=0$, and equations (45) to (47) reduce to

$$w_m=\frac{\beta}{\sqrt{\pi x}}+\frac{\beta^2}{2\pi x}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

where

$$x=\tfrac{1}{4}(x_1/\mathrm{L}+c). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

From fig. 10 we see that K($=4\beta$) tends to about 1·86 at the rear edge of the model. In order to preserve the analogy with the treatment of the flat plate wakes, we shall take $4\beta=1{\cdot}86$.

Equation (52) then becomes

$$w_m=\frac{0{\cdot}2622}{\sqrt{x}}+\frac{0{\cdot}0344}{x} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

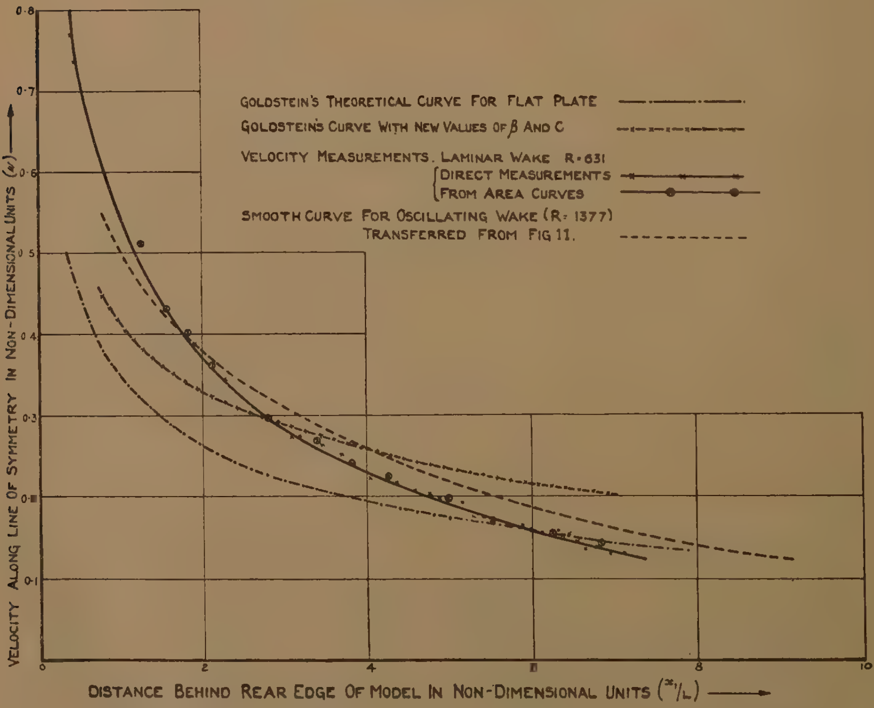
The value of c may be so chosen that the calculated and experimental values of the velocity along the line of symmetry agree at any specified point. Table V. exhibits the results obtained.

TABLE V.

Effect of Choice of Origin on Theoretical Velocity Field.

Value of x at which results agree.	Value of w_m (eq. (54)).	Corresponding value of x_1/L (fig. 9).	Value of c (eq. (53)).
0·5	0·4398	1·5	0·5
1	·2966	2·8	1·2
2	·2027	4·7	3·3
3	·1628	5·8	6·2

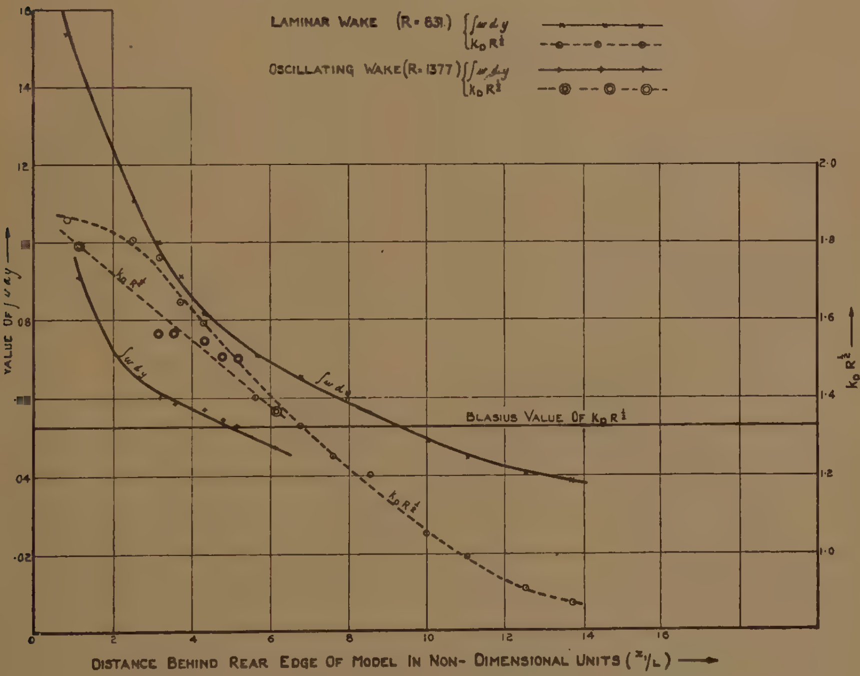
Fig. 9.



Velocity along the line of symmetry of the wake. Aerofoil section.

Coincidence of the calculated and experimental velocities at $x=1$ ($x_1/L=2{\cdot}8$) was selected, and hence c was taken as 1·2. The resulting curve for w_m is shown in fig. 9, and it is seen that the agreement is much less close than for the flat plate cases. The wake is about three times as wide as that given by the flat plate, and hence the effect of the walls of the tank is considerably greater. Inspection of figs. 5 and 9 suggests that the increase of the slope of the experimental velocity curve as compared with the calculated curve is due to the effect of the walls of the tank.

Fig. 10.



Variation of $\int w\,dy$ and $k_D R^{\frac12}$ with x_1/L. Aerofoil section.

Fig. 10 exhibits the same general features as fig. 8 (flat plate) except that the variation of K with x_1/L is greater and the results are more irregular. The general shape of the velocity profile curves is similar to those previously discussed, and comparisons with Goldstein's calculations and the Gaussian error curve have not been made.

16.2. *Oscillating Wake.* R=1377.

A sequence of four photographs of the wake were obtained, of which the first three are shown in Pls. XI.–XIII. The wake began to oscillate at a distance of about 30 cm. behind the model (second plate). We have:

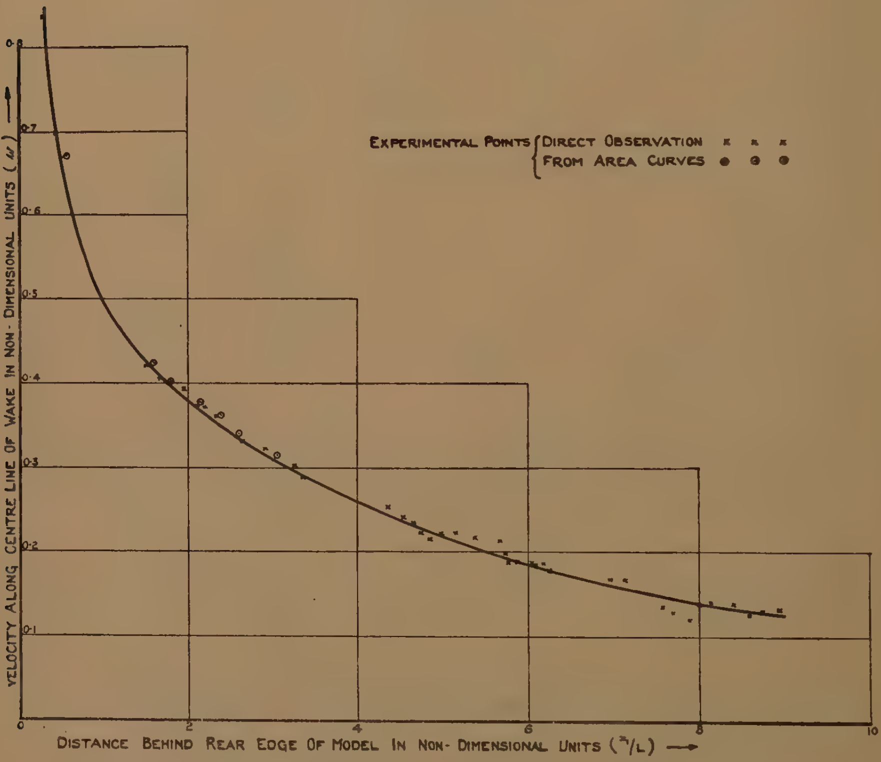
Velocity of model...... $U_0=1{\cdot}487$ cm./sec.

and

Chord of model........ $L=10{\cdot}16$ cm.

An attempt, for reasons which will emerge in § 23, was made to measure the velocity along the centre line of the wake even after the oscillation had commenced. The centre of the wake was estimated and lengths of the particle tracks were measured in the local direction of motion of the fluid.

Fig. 11.



Velocity along the centre line of the wake. Aerofoil section. Oscillating Wake.

The observations are plotted in fig. 11, and the oscillating nature of the results is very evident in the range $x_1/L>4$. A smooth curve was drawn through the observations, and transferred to fig. 9 for comparison with the laminar wake.

Velocity measurements along lines perpendicular to the centre line of the wake could only be made on the first two plates, before the oscillation was appreciable.

The values of $\int w\,dy$ and $k_D R^{\frac{1}{2}}$ were calculated in the usual way, the results being plotted in fig. 10 and tabulated in Table VI.

TABLE VI.

Summary of Area and Drag Calculations for Oscillating Aerofoil Wake.

Plate.	x_1/L.	w_m.	$\int w\,dy$.	K $=k_D R^{\frac{1}{2}}$.
1	0·55	0·670	0·0918	1·79
2	1·57	·425	·0604	1·57
	1·78	·400	·0592	1·57
	2·12	·380	·0573	1·55
	2·38	·362	·0548	1·51
	2·59	·340	·0531	1·50
	3·05	·315	·0475	1·36

These results for the oscillating wake are, of course, of a lower degree of accuracy than those obtained when the wakes are laminar.

A comparison of the curves for the two plates indicates that K($=k_D R^{\frac{1}{2}}$) is almost independent of the Reynolds number, but exhibits considerable variation with x_1/L.

In general we may say that the wakes produced by the aerofoil section are substantially similar to those produced by the flat plate, the main differences being attributable to the greater effect of the walls of the tank in the former case.

17. *Comparison of Theory and Experiment for the Oscillating Wake.*

The measurements of laminar wakes have shown that their velocity distribution approximates fairly closely to a one-dimensional profile of a Gaussian error law form; such a velocity profile has been considered in Part I. Since the wakes are observed to exhibit sinusoidal oscillations when an appropriate Reynolds number is reached, we may attempt a comparison between the theory of Part I. and measured characteristics of experimental wakes.

It may be noted that the theory provides an exact evaluation of the wave-velocity, whereas the wave-length of the oscillations is only derived by an approximate method. An intrinsic deficiency of the theory is its inability to predict any critical Reynolds number for instability.

18. *The Observed Oscillating Wake.*

The whole sinusoidal wave-form was observed to move slowly in the direction of motion of the body. When a sufficiently large Reynolds number is reached the oscillating wake breaks up at some distance downstream into isolated vortices; and these vortices were observed to move with the same velocity as that of the sinusoidal wave-form produced nearer the model.

It was further noted that the velocity either of the oscillating wave form or of the discrete vortices, when expressed in non-dimensional units (as a fraction of the model velocity), did not vary greatly with the Reynolds number. A summary of the results of various experimenters is given by Richards *, and the general statement is made that "the smoother the form of the body producing the wake, the smaller is the velocity of the vortices relative to the fluid in which they are formed." (The velocity here referred to is expressed in non-dimensional units as above.)

The mean values here obtained are 0·090 for the flat plate and 0·068 for the aerofoil section, harmonizing in a general manner with the result of previous workers.

The method of measuring the velocity of the wave-form is similar to that adopted by Richards †. Particular crests and troughs of the oscillation could be identified in successive photographs, and the velocity obtained by "superposing" the plates and measuring the distance the wave had travelled in the known time between the two plate-exposures.

Four series of photographs of the oscillating wake were obtained, and the relevant data are presented in Table VII.

The results are inadequate to provide any general conclusions as to the variation of λ and c with the Reynolds number, but are in general agreement with Richards' results ‡. The wakes photographed in Plates F. 21–24 and F. 9–12 correspond to almost equal Reynolds numbers. The values of the wave velocity, c, for the two sets of plates are in agreement, but there is a marked difference in the wave-length. The oscillations on Plates F. 9–12 began about 8 cm. behind the rear edge of the plate, while those on Plates F. 21–24 began much nearer.

---

* Aero. Res. Comm. Reports and Memo. no. 1590 (1933).
† *Loc. cit.*
‡ *Loc. cit.*

TABLE VII.

Particulars of Measured Oscillating Wakes.

Ref. nos. of plates.	U_0 (cm./sec.).	Reynolds number, R.	Wave-velocity c (cm./sec.).	$\frac{c}{U_0}$.	Wave-length λ (cm.).	$\frac{\lambda}{L}$.
		Flat plate (L=3·81 cm.).				
F. 17–20 .	2·49	791	0·197	0·079	3·30	0·687
F. 21–24 .	3·06	1110	·291	·095	2·75	·722
F. 9–12 .	3·06	1090	·291	·095	3·07	·806
		Aerofoil section (L=10·16 cm.).				
—	1·49	1377	·102	·068	8·40	·827

19. *Correlation of Theoretical and Experimental Units.*

We shall now, for the sake of clarity, add a suffix, " 1," to those quantities which are expressed in terms of the non-dimensional units of Part I. The notation of § 12 will be retained.

The results which are relevant to the comparison are then expressed by the following equations :—

$$U_1(y_1)=e^{\frac{1-y_1^2}{2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

$$\int_0^\infty U_1(y_1)\,dy_1=2{\cdot}065, \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

$$c_1=\left|\,y_1\,\right|_{U_1''(y_1)=0}=1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

The value of α_1 will be considered to lie between the limits 1·25 and 1·45 (see Table I.).

U_m and h are the representative units of velocity and length respectively (as defined in Part I.) for the profile given in equation (55). The Reynolds number, R_m, of this profile has been defined by the equation

$$\nu R_m=U_m h. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

We shall also, as in § 15, write w_m for the value of w along the line of symmetry of the wake.

The connexions between the theoretical and experimental quantities are therefore given by

$$\frac{\int_{-\infty}^{\infty} W dy}{U_m h}=\frac{U_0 L}{U_m h}\int_{-\infty}^{\infty} w.dy=4{\cdot}130, \quad . \quad . \quad . \quad . \quad (59)$$

$$\frac{w_m U_0}{U_m} = 1{\cdot}649, \qquad \qquad (60)$$

$$\frac{c}{U_m} = c_1 = 1, \qquad \qquad (61)$$

$$\left.\begin{aligned} \frac{\lambda}{h} = \lambda_1 &= \frac{2\pi}{1{\cdot}25}\,\text{(upper limit)}, \\ &= \frac{2\pi}{1{\cdot}45}\,\text{(lower limit)}, \end{aligned}\right\} \qquad \qquad (62)$$

where λ and c are respectively the measured wave-length and wave-velocity in dimensional units. In order to connect the non-dimensional quantities of theory with the dimensional quantities of experiment, it is necessary to determine the magnitude of U_m and h, which involves measurement of the maximum velocity of, and the area under, the experimental velocity profile curve.

The shape and size of the velocity profile of the observed wake depend on the distance behind the model, so U_m, h, and R_m will vary downstream. Instability sets in at a particular point behind the model, and the resulting oscillation persists downstream with the wave-length and wave-velocity appropriate to this point. The values of U_m and h which must be taken are therefore those appropriate to that point of the profile at which instability sets in. The determination of these quantities presents certain points of difficulty, since while c and λ must be obtained from photographs of an oscillating wake, U_m and h must be determined from photographs of a wake before oscillation has begun. Since U_m and h vary downstream, some degree of extrapolation is necessary.

20. *Measurement of the Representative Velocity and Length.*

The non-dimensional velocity, w_m, has been shown (§ 14) to vary with the distance behind the model and also, to some extent, with R.

Figs. 5 and 9 provide the means of estimating w_m for any given value of x_1/L. The value of w_m is not very sensitive to changes of Reynolds number, and a sufficiently accurate interpolation can be made for any desired value of R. The evaluation of U_m then follows immediately from equation (60).

The determination of h must be more indirect, since the value of $\int_{-\infty}^{\infty} w\,dy$ depends both on x_1/L and R, and the experimental material is not adequate to justify extrapolation to the higher Reynolds numbers of the oscillating wakes.

Figs. 8 and 10 indicate, however, that $k_D R^{\frac{1}{2}}$ (denoted by K) is approximately independent of R and depends only on x_1/L.

Equation (49) gives

$$k_D=\int_{-\infty}^{\infty}(w-w^2)\,dy, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

from which $\int_{-\infty}^{\infty} w\,dy$ has been obtained by assuming that $\int_{-\infty}^{\infty} w^2\,dy \Big/ \int_{-\infty}^{\infty} w\,dy$ has the same value for the actual wake and for the equivalent Gaussian profile. The evaluation of $\int_{-\infty}^{\infty} w\,dy$ thus involves estimates of w_m and K, both of which have been shown to be independent of R. The accuracy of this method is exhibited by Table VIII.

TABLE VIII.

Accuracy of Approximation used in deriving the Representative Length of the Profile.

Plates.	x_1/L.	$\int_{-\infty}^{\infty} w^2\,dy$ as per cent. of $\int_{-\infty}^{\infty} w\,dy$.	
		Calculated from equivalent Gaussian profile.	By analysis of experimental observations.
F. 5–7	2·59	22·6	22·3
	7·93	11·3	13·3
	9·27	9·2	10·3
F. 13–16	6·29	14·1	14·8
	7·83	12·0	13·0
	8·75	11·3	11·7
	12·00	9·2	10·0
	18·26	7·07	6·75

21. *Difficulties of the Comparison.*

It has already been remarked that the characteristics of the oscillating wake are determined by the point at which instability first sets in ; hence the correct estimation of this point is of the greatest importance. The matter is complicated by the existence of what have been termed "pseudo-oscillations," which are largely conditioned by small residual motions of the water in the tank. The oscillations in each of the flat plate wakes for which comparison is made are of this nature.

The basic result of Part I. of this paper can be expressed in the form—the wave-velocity of the oscillations of the wake (considered infinitesimally small) is equal to the velocity at the point of inflexion of the laminar velocity profile on which the small oscillations may be considered to be superposed.

The difficulty of estimating the position of the point of inflexion for an experimental profile is clearly brought out by a consideration of figs. 6 and 7, which show typical profiles. The shape of these profiles near the edges of the wake tends to displace the point of inflexion somewhat further from the line of symmetry than is the case for the " Gaussian " profile. After inspection of all the measured velocity profiles, it was decided to assume that the velocity at the point of inflexion was one-half of the maximum velocity of the profile. Equation (60), which may be considered as defining the representative velocity U_m, was accordingly replaced by

$$U_m=\tfrac{1}{2}\,w_m U_0, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (64)$$

the other equations remaining unchanged.

The three main difficulties encountered in making the comparison of the theoretical and experimental results may be summarized thus :—

(1) Estimation of the point where instability sets in.

(2) Extrapolation of $\int_{-\infty}^{\infty} w\,dy$ and w_m from laminar to oscillating wakes.

(3) Estimation of the points of inflexion of the experimental velocity profiles.

22. *Comparison for Wakes produced by a Flat Plate.*

Comparisons were made for three oscillating wakes produced by the flat plate.

22.1. *F. 17–20.* R=791.

$$U_0=2{\cdot}49 \text{ cm./sec.}$$

$$L=3{\cdot}81 \text{ cm.}$$

The wake exhibited small " pseudo-oscillations " which were sufficiently marked for measurement only on the second and third plates (F. 18 and 19). The wave-length, λ, remained sensibly constant at 3·3 cm., and the wave-velocity, c, was measured as 0·197 cm./sec., giving $c/U_0=0{\cdot}079$.

The oscillations were estimated to begin in the neighbourhood of $x_1/L=8$.

Table IX. shows the results of the comparison and also indicates the effect of varying the value of x_1/L assumed for the beginning of the oscillation.

TABLE IX.

Comparison of Theory and Experiment for Flat Plate Wake R=791.

	Estimated x_1/L at which oscillation begins.	w_m.	$K = k\ R^{\frac{1}{2}}$.	c (cm./sec.).	$\int_{-\infty}^{\infty} w\,dy$.	h (cm.).	λ (cm.) (upper and lower limits).
Theory	8·0	0·17	1·32	0·212	0·0532	0·578	2·90–2·50
	8·5	0·16	1·30	0·199	0·0521	0·601	3·02–2·605
	10·0	0·14	1·22	0·174	0·0481	0·635	3·19–2·75
Experiment.	—	—	—	0·197	—	—	3·3

22.2. *F. 21–24.* R=1110.

$$U_0 \doteq 3{\cdot}06 \text{ cm./sec.}$$

$$L = 3{\cdot}81 \text{ cm.}$$

The oscillations were very marked, the maximum amplitude being about ½ cm. They commenced about 6 cm. behind the rear edge of the plate (corresponding to x_1/L about 1·5), but the amplitude varied considerably and almost vanished at points on the third and fourth plates,

TABLE X.

Comparison of Theory and Experiment for Flat Plate Wake R=1110.

	Assumed x_1/L at which oscillation begins.	w_m.	$K = k_D R^{\frac{1}{2}}$.	c (cm./sec.).	$\int_{-\infty}^{\infty} w\,dy$.	h (cm.).	λ (cm.) (upper and lower limits).
Theory	1·5	·48	1·65	0·735	·0751	·288	1·46–1·26
	4	·26	1·50	0·3975	·0552	·392	1·99–1·71
	6	·19	1·40	0·291	·0486	·471	2·39–2·06
Experiment.	—	—	—	0·292	—	—	2·75

indicating the effects of small residual motions of the water. The measured wave-length varied between 2·6 and 2·9 cm.; a mean value of 2·75 cm. was taken.

The measured wave-velocity, c, was 0·292 cm., giving $c/U_0=0{\cdot}095$.

Comparison was first made taking $x_1/L=1{\cdot}5$ as the point at which the oscillation began. Large discrepancies were thus obtained, and somewhat better agreement resulted from taking larger values of x_1/L, as shown in Table X.

22.3. *F. 9–12.* R=1090.

$$U_0=3{\cdot}065 \text{ cm./sec.}$$
$$L=3{\cdot}81 \text{ cm.}$$

The mean measured wave-length, λ, was 3·07 cm. and the wave-velocity, c, was 0·292 cm./sec., as in § 22.2.

The calculations were not made in this case, as they are practically identical with those of § 22.2.

23. *Comparison for Wake produced by an Aerofoil Section.*

The comparison was made for the oscillating wake previously measured in §16·2, corresponding to a Reynolds number of 1377. General data for this wake have been given in §16·2 and Table VII.

The oscillations were estimated to begin at about $x_1/L=3$, and the first calculations were made using this value. Large discrepancies were obtained between the experimental and theoretical values of both the wave-length and wave-velocity. Larger values of x_1/L were assumed, but considerable discrepancies still remained, as indicated in Table XI.

TABLE XI.

Comparison of Theory and Experiment for
Aerofoil Wake R=1377.

	x/L.	w_m.	$\int_{-\infty}^{\infty} w\,dy$.	c.	h.	λ.
Theory ……	3	0·310	0·0490	0·230	0·781	3·93–3·385
	7·5	0·150	0·0256	0·112	0·850	4·28–3·685
Experiment…	—	—	—	0·102	—	8·4

The absence of the theoretical results which were employed in the flat plate calculations necessitated a closer reliance on experiment. For this reason an attempt was made in §16·2 to obtain a “ fictitious ” value

General Data.

Flat plate. $1\frac{1}{2}''$ chord.
Aerofoil section. $4\frac{1}{2}''$ chord. Symmetrical section "R.A.F. 30."
Zero angle of incidence.

Flat plate. "F. 5." $U_0=1{\cdot}905$ cm./sec.
$R=636$.

Laminar wake just behind plate.

of w_{m} by direct measurement in the range of oscillation of the wakes. It was not possible to do this for $\int_{-\infty}^{\infty} w\,dy$ when $x_1/\mathrm{L}>3$. The results shown on fig. 10 indicate that K is approximately independent of the Reynolds number, and hence $\int_{-\infty}^{\infty} w\,dy$ may be calculated by the indirect method already used for the flat plate wakes, making use of the values of K obtained for the laminar wake ($\mathrm{R}=631$) for $x_1/\mathrm{L}>3$.

24. *The Influence of the Tank Walls.*

The theoretical treatment of Part I. applies to an infinite field of fluid, whereas the wakes on which measurements were made were obtained in a water tank of finite size. The general effect of the tank walls is of a retarding nature; fig. 5 shows that the value of w falls off with increase of x_1/L at a somewhat greater rate than is predicted by theory. The effect of the tank-walls in reducing the value of $\mathrm{K}(k_{\mathrm{D}}\mathrm{R}^{\frac{1}{2}})$ downstream is shown in figs. 8 and 10; the reduction is, as anticipated, more marked with the aerofoil wake. The work of Fage * and others suggests that the effect of the walls of the tank can be allowed for by assuming that the value of K remains constant at its value when $x_1/\mathrm{L}=0$.

This value of K was taken as 1·600 (25 per cent. in excess of Blasius's value) for the flat plate and as 1·86 (from fig. 10) for the aerofoil section.

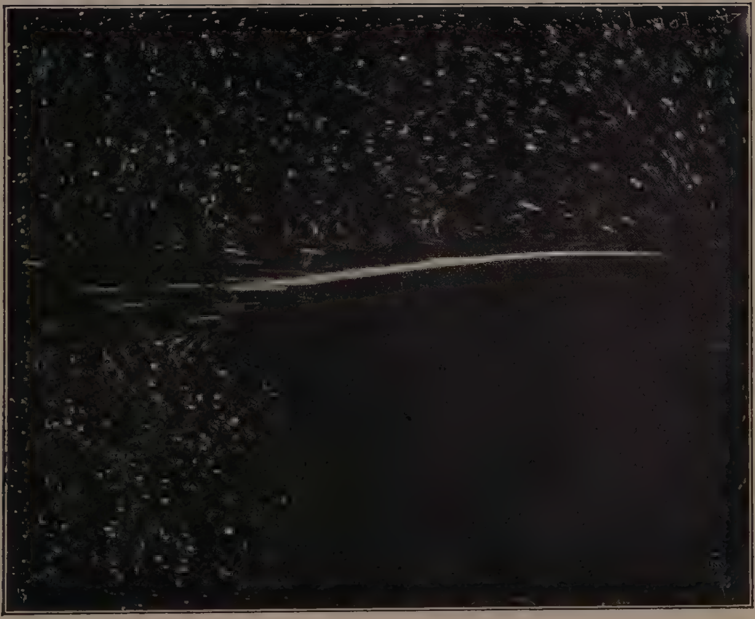
TABLE XII.

Comparison allowing for Effect of the Walls of the Tank.

R.	x_1/L.	w_{m}.	Calculated.		Measured.	
			c.	λ.	c.	λ.
Flat plate. $k_{\mathrm{D}}\mathrm{R}^{\frac{1}{2}}=1\ 66$.						
791	8·5	0·16	0·199	3·89–3·36	0·197	3·30
1110	6·0	0·19	0·291	2·81–2·42	0·292	2·75
Aerofoil. $k_{\mathrm{D}}\mathrm{R}^{\frac{1}{2}}=1\cdot86$.						
1377	7·5	0·15	0·112	9·24–7·96	0·102	8·4

---

* *Loc. cit.*



Aerofoil section. $U_0 = 1{\cdot}487$ cm./sec.
$R = 1377$.

Photograph showing development of the oscillation of the wake.
Exposure about 1 second.

Aerofoil section. $U_0 = 1{\cdot}487$ cm./sec.
$R = 1377$.

Photograph showing development of the oscillation of the wake.
Exposure about 1 second.

Centre of wake about 12 in. behind model.

Aerofoil section. $U_0=1{\cdot}487$ cm./sec.
$R=1377$.

Photograph showing development of the oscillation of the wake.
Exposure about 1 second.

Centre of wake about 21 in. behind model.

26. *Summary.*

Part I.—The stability of certain types of "one-dimensional" steady flow in an unlimited viscous fluid has been investigated by the method of small oscillations. The general result is obtained that the velocity of the oscillations (in the direction of motion of the main flow) is equal to the velocity of the main flow at the point of inflexion of the laminar velocity profile.

Two profiles of physical interest are considered in greater detail and approximate values obtained for the wave-length of the oscillations. No critical Reynolds number for the transition to instability can be obtained by the method.

Part II.—Photographs have been obtained of both laminar and oscillating wakes produced by models (flat plates and symmetrical aerofoil sections at zero incidence) moving in a tank of water. Oscillations do not occur when the Reynolds number is less than 600 for the flat plate and 1000 for the aerofoil section, and it is probable that the onset of instability can be considerably delayed by careful elimination of all "residual motions" in the water.

The velocity-field in the laminar wake has been measured, and a comparison made with certain results of Goldstein and Fage. Good agreement is obtained, the divergence being appreciable at the edges of the wake, due to the effect of the walls of the tank.

Measurements of the wave-length and wave-velocity have been made for various oscillating wakes, and a comparison with the conclusions of Part I. attempted. The agreement between theory and experiment is fairly satisfactory, but the comparison is blunted owing to several difficulties of interpretation of the theoretical results.

My thanks are due to Professor L. Bairstow, C.B.E., A.R.C.Sc., F.R.S., for his direction of the work; to the Department of Scientific and Industrial Research for a grant which made the research possible; to H. B. Squire, Esq., M.A., for valuable suggestions and criticisms of the theoretical part of the paper; and to various members of the Aeronautical Department of the Imperial College of Science and Technology for assistance in the experimental work.

XXVI. *The Principles of a Mechanical Method for Calculating Regression Equations and Multiple Correlation Coefficients and for the Solution of Simultaneous Linear Equations.*

By T. E. W. Schumann, M.Sc., Ph.D.*

[Received March 4, 1939.]

1. *Introductory.*

During the last two or three decades great advances have been made in the theory of statistics, in great measure due to the fact that this division of mathematics is playing an increasingly important rôle in biometrics, economics, agriculture, meteorology, biology, etc., with the consequent demand for the greater elaboration of the theory. This statement applies particularly to that part of statistical theory dealing with correlation and equations of regression.

Now, the methods of determining linear regression equations and calculating simple and multiple correlation coefficients are well known, and are adequately dealt with in a number of text-books; but when it comes to the actual carrying out of the numerical work, especially where a large number of variables is involved, the task becomes so overwhelming that one is often inclined to retreat in dismay.

It is, therefore, only natural that ways and means should be sought for reducing the volume of numerical work; and the mechanical model, to be described presently, was conceived in an effort to find some means of eliminating the more laborious part of the arithmetical work when multiple correlation coefficients are being dealt with.

In order clearly to demonstrate the amount of labour involved, and also to serve as a concrete basis for further discussion, we present here a very brief review of the essentials of linear regression equations and of multiple correlation.

2. *Theory of Correlation: Two Independent Variables.*

For the sake of simplicity and in order to avoid a rather complicated notation, we consider the case of one dependent variable $X_1, X_2 \ldots . X_n$, and of only two independents $Y_1, Y_2 \ldots . Y_n$ and $Z_1, Z_2 \ldots . Z_n$.

If we write small letters instead of capitals for the *departures* of these variables form their respective average values, then the equations of linear regression may be written

---

* Communicated by the Author.

$$\left.\begin{array}{l} x_1=k_2y_1+k_3z_1+d_1, \\ x_2=k_2y_2+k_3z_2+d_2, \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ x_n=k_2y_n+k_3z_n+d_n, \end{array}\right\} \quad \ldots\ldots\ldots \quad (1)$$

where, for each set of values, there is a certain residual $d_1d_2 \ldots . d_n$.

In determining the values of the constants k_2 and k_3 the requirement is that the sum of the squares of the residuals, *i. e.*, of Σd^2, becomes a minimum.

But $\Sigma d^2=\Sigma(x-k_2y-k_3z)^2$, and by differentiating this equation first with respect to k_2, then with respect of k_3, and equating to zero in each case, we find that

$$\Sigma xy=k_2\Sigma y^2+k_3\Sigma yz, \quad \ldots\ldots\ldots \quad (2)$$

$$\Sigma zx=k_2\Sigma yz+k_3\Sigma z^2, \quad \ldots\ldots\ldots \quad (3)$$

k_2 and k_3 being called the partial regression coefficients of x on y and z.

The multiple correlation coefficient R may now be defined as the extent to which the variations of X are determined by those of Y and Z, *i. e.*, R is the mean value of k_2y+k_3z divided by the mean value of x, the definition of a mean in this case being the square root of the mean sum of the squares. Hence

$$\begin{aligned} \mathrm{R}^2 &= \Sigma\frac{(k_2y+k_3z)^2}{\Sigma x^2} \\ &= \frac{k_2(k_2\Sigma y^2+k_3\Sigma yz)+k_3(k_2\Sigma yz+k_3\Sigma z^2)}{\Sigma x^2} \\ &= \frac{k_2\Sigma xy+k_3\Sigma xz}{\Sigma x^2} \quad \ldots\ldots\ldots\ldots \quad (4) \end{aligned}$$

The final expression follows immediately from (2) and (3).

It will be seen below that the proposed machine allows of a direct determination of k_2 and k_3, and also of Σxy and Σxz. Hence the multiple correlation coefficient R can be readily determined with the aid of equation (4).

We may now introduce the conventional symbols

$$\left.\begin{array}{l} s_1{}^2=\Sigma x^2, \\ s_2{}^2=\Sigma y^2, \\ s_3{}^2=\Sigma z^2, \end{array}\right\} \quad \ldots\ldots\ldots$$

$$\left.\begin{array}{l} r_{12}=\Sigma xy/s_1s_2, \\ r_{23}=\Sigma yz/s_2s_3, \\ r_{13}=\Sigma zx/s_1s_3, \end{array}\right\} \quad \ldots\ldots\ldots \quad (6)$$

in which case (4) reduces to

$$R^2 s_1 = k_2 s_2 r_{12} + k_3 s_3 r_{13}. \quad \ldots\ldots \quad (7)$$

The results obtained for three variables may readily be extended to any number of variables m. In this case we obtain $m-1$ linear equations corresponding to (2) and (3) above :—

$$\left.\begin{array}{l} r_{12}s_1 = k_2 s_2 + k_3 s_3 r_{23} + \ldots\ldots \quad k_m s_m r_{2m}, \\ r_{13}s_1 = k_2 s_2 r_{23} + k_3 s_3 + \ldots\ldots \quad k_m s_m r_{3m}, \\ r_{14}s_1 = \text{etc.} \end{array}\right\}, \quad \ldots\ldots \quad (8)$$

and the solutions for $k_2 \ldots\ldots k_m$ may be written down in determinant form in the usual way. In analogy with (7) we find that

$$R^2 s_1 = k_2 s_2 r_{12} + k_3 s_3 r_{13} + \ldots\ldots k_m s_m r_{1m} \quad \ldots\ldots \quad (9)$$

The value of R may also be expressed in determinant form, thus

$$(1-R^2)\begin{vmatrix} 1 & r_{23} & r_{24} & \ldots\ldots & r_{2m} \\ r_{23} & 1 & r_{34} & \ldots\ldots & r_{3m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{2m} & r_{3m} & \cdot & \ldots & 1 \end{vmatrix} = \begin{vmatrix} 1 & r_{12} & r_{13} & \ldots\ldots & r_{1m} \\ r_{12} & 1 & r_{23} & \ldots\ldots & r_{2m} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ r_{1m} & r_{2m} & \cdot & \ldots & 1 \end{vmatrix} \quad \ldots \quad (9a)$$

the derivation of which equation may be found in any text-book on correlation.

Formally the problem of obtaining the constants of the regression equation and the value of the correlation coefficient R is thus solved, but in many practical cases the numerical evaluation is a long and tedious process, as will now be demonstrated.

3. *The Labour of Numerical Calculation.*

In an attempt to estimate the time required to calculate multiple correlation coefficients, let us consider the concrete case of nine independent variables, each series consisting of 25 items, *i. e.*,

$$m = 10$$

and

$$n = 25.$$

Before the multiple correlation coefficient R can be determined it is first necessary to calculate all the coefficients r_{pq} between each two sets of variables, in all amounting to $m\dfrac{(m-1)}{2}$, or 45 coefficients when $m=10$. Experience shows that the calculation and necessary checking occupies about 10 minutes for each coefficient. Hence the total time consumed here is roughly $5m^2$ minutes.

Next, in solving the nine equations (8), one may either evaluate the determinants or employ one of several other short methods, Whatever method is employed, however, the number of distinct operations of addition, subtraction, multiplication or division is roughly proportional to the third power of m. Using the Doolittle method, which is, perhaps, as short as any other, it was found that the time roughly amounted to $\frac{m^3}{4}$ minutes for solving the $m-1$ simultaneous equations, a slide-rule being used and great accuracy thus sacrificed.

Hence, if t_1 denotes the time in minutes for the complete calculation, then

$$t_1 = 5m^2 + \tfrac{1}{4}m^3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

is probably a fair estimate. On this basis it would take an (experienced) worker 750 minutes or nearly two working days to calculate one multiple correlation coefficient involving nine independent variables.

These considerations show that a shorter and more effective method of dealing with problems of this nature is very much to be desired, and in the following paragraphs a mechanical method of eliminating the lengthy numerical calculations will be described.

4. *Description of the Mechanical Apparatus.*

In order to make the description of the apparatus as simple as possible, we confine ourselves to the case of two independent variables, but the theory applies equally well to any number of independents.

The figure (p. 262) is a schematic drawing of the mechanical model, in which B_2 represents a beam pivoted at its extremities and free to rotate about its horizontal axis. To the beam weightless arms are rigidly attached, all lying in the same horizontal plane. The lengths of these arms are proportional to the departures of the first independent variable $y_1\ y_2 \ldots . y_n$, those to the right of the axis representing positive values and the rest, negative.

A similar beam B_3 carries weightless arms proportional to $z_1\ z_2 \ldots . z_n$.

A weightless inextensible string is attached to the hook E_1, passes downwards underneath the small pulley situated at y_1, back over the pulley P_1, down again to the pulley at z_1 on the first arm of the second beam, and thence over the pulleys p_1 and q_1. Attached to the end of the string is an iron rod F_1, which hangs partly submerged in the mercury contained in the trough HH.

A second inextensible string is attached at the point E_2 and passes in turn round the pulleys at y_2, P_2, z_2, p_2, and q_2, and terminates in the iron rod F_2. In the same way a third string acting on beams B_2 and B_3 at the points y_3 and z_3 terminates in the iron rod F_3, and so on. On the diagram only four rods, F_1 to F_4, are shown, but the number can be extended indefinitely.

Initially the rods, which are all of equal weight and size, are adjusted to hang at the same level in the mercury, the beams B_2 and B_3 being in their "normal" position such that their arms are all horizontal. In order to ensure that the strings are always under tension weights may be added to F_1, F_2 etc., provided that the same weight is attached to each string.

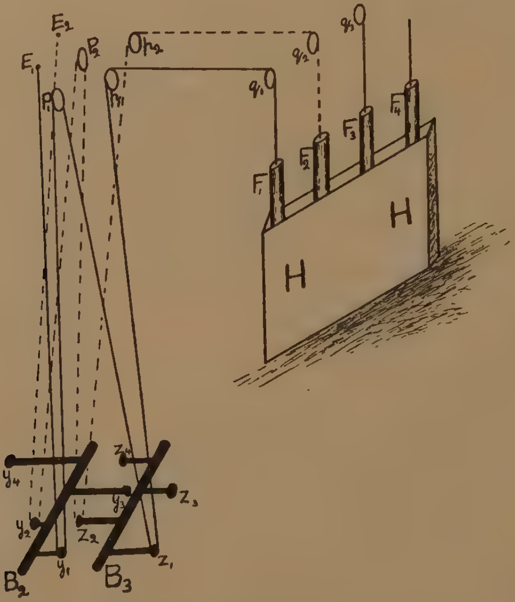


Illustration of the mechanical principle.

Furthermore, we assume that the pulleys $P_1p_1 \ldots . P_np_n$ are at a considerable distance above the beams, so that the strings may all be considered to be essentially vertical.

If the proper adjustment is thus made, it is evident that the tensions in the strings are all equal and that, consequently, the sum of the moments of the forces acting either on B_2 or B_3 is equal to zero. This immediately follows from the fact that $y_1 \ldots . y_n$, $z_1 \ldots . z_n$ represent *departures* from the average values and that, consequently, $\Sigma y = \Sigma z = 0$.

The hooks $E_1 \ldots . E_n$, to which the strings are attached, are made to be movable, and we now want to investigate what happens when these hooks are moved through vertical distances proportional to $x_1 \ldots . x_n$, in which the upward direction is regarded as positive. Clearly, since the total length of the string connecting E_1 to F_1 remains unaltered, the movement of E_1 will be communicated to the rod F_1, whose effective weight will be altered due to the alteration in its volume submerged in the mercury. The tension in the first string will thus be altered by the displacement of E_1, and in the same way the tensions in the other strings will also be altered. At the start, when all the rods hang at the same level and the tensions in the strings are all equal, the moments of the forces on the two beams are zero and they are in (neutral) equilibrium, but by the displacements specified above the tensions in the strings are altered and the equilibrium will be disturbed. If the beams are free to rotate they will now come to rest in new positions of equilibrium on the supposition that frictional forces are absent.

Suppose that when equilibrium has been attained the beam B_2 has rotated through an angle β_2 in an anticlockwise direction, and B_3 through an angle β_3. From the geometry of the system, it is evident that the pulley at y_1 moves through an upward distance $y_1 \sin \beta_2$ and the pulley at z_1 through a distance $z_1 \sin \beta_3$. Due to these two rotations the iron rod F_1 therefore moves down a distance w_1 where

$$w_1 = 2y_1 \sin \beta_2 + 2z_1 \sin \beta_3. \quad \ldots \ldots \quad (11)$$

Hence the total upward displacement of the rod F_1 due to the displacement x_1 of E_1 and to the two rotations is $x_1 - w_1$. Now it may readily be proved that, when the movements of all the rods are taken into account, the level of the mercury in the trough remains unchanged; and hence, if T_1 denotes the increase in tension of the first string, it is equal to the alteration in volume of the mercury displaced by F_1 multiplied by its density and by g, i. e.,

$$T_1 = a\rho g(x_1 - w_1),$$

where a = cross-sectional area of the rod,
ρ = density of mercury,
g = acceleration due to gravity.

By similar reasoning it is found that the change in tension T_f of the fth string is

$$\begin{aligned} T_f &= a\rho g(x_f - w_f) \\ &= a\rho g(x_f - 2y_f \sin \beta_2 - 2z_f \sin \beta_3), \quad \ldots \ldots \quad (12) \end{aligned}$$

and the argument applies equally well whether the value of $(x-w)$ or T is positive or negative.

By considering the equilibrium of the beams B_2 and B_3 in turn we are now in a position to calculate the values of β_2 and β_3 in terms of known quantities. Consider the moments of the forces acting on B_2. The first string exerts a force $2T_1$ at y_1 and the moment about the axis of the beam is $2T_1y_1 \cos \beta_2$. Similarly, the second string exerts a moment $2T_2y_2 \cos \beta_2$, and hence the sum of the moments is 2 cos $\beta_2 \Sigma \mathrm{T}y$, and this must be equal to zero, since the beam is in equilibrium, *i. e.*,

$$\Sigma \mathrm{T}y=0.$$

The beam B_3 is also in equilibrium, and hence

$$\Sigma \mathrm{T}z=0.$$

Substituting the values of T from (12), we find that

$$\begin{aligned} \Sigma xy-2 \sin \beta_2 \Sigma y^2-2 \sin \beta_3 \Sigma yz=0, \\ \Sigma xz-2 \sin \beta_2 \Sigma yz-2 \sin \beta_3 \Sigma z^2=0, \end{aligned}$$

or, making use of the notation introduced previously,

$$\left. \begin{aligned} r_{12}s_1=2s_2 \sin \beta_2+2s_3 r_{23} \sin \beta_3, \\ r_{13}s_1=2s_2 r_{23} \sin \beta_2+2s_3 \sin \beta_3, \end{aligned} \right\} \quad . \; . \; . \; . \; . \quad (13)$$

equations which are identical with (2) provided

$$\left. \begin{aligned} k_2=2 \sin \beta_2, \\ k_3=2 \sin \beta_3. \end{aligned} \right\} \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (14)$$

With the aid of this mechanical device we are, therefore, able to determine the values of the constants k_2 and k_3 of the regression equation simply by reading off the values of β_2 and β_3 on suitable scales, and looking up the sines of these angles.

It is perfectly clear that the appliance would work in the same way if the number of rotating beams were increased. Thus, if we were dealing with $m-1$ independent variables, we would require $m-1$ beams, with a corresponding increase in the sets of overhead pulleys, and the values of k_2 to k_m would be equal to double the sines of the angles of deflexion of the respective beams.

Having determined the values of the k's, our final step is to calculate the multiple correlation coefficient R. This can easily be accomplished by taking a further set of readings of the angles $\gamma_2 \ldots . \gamma_m$, where these angles are obtained in the following way. Keep all the beams, except the first, fixed in their " normal " horizontal position. Let γ_2 be the angle of deflexion of the first beam caused by the displacement of the points of suspension $\mathrm{E}_1 \ldots . \mathrm{E}_n$ through distances $x_1 \ldots . x_n$. Then, since the angles of deflexion of all the other beams are each equal to zero, it follows from (13) that

$$r_{12}s_1=2\, s_2 \sin \gamma_2 \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (15)$$

with corresponding equations for the other beams when each of them in turn is left free to rotate, the rest being fixed in their " normal " position.

Hence, substitution of $k_1 \ldots . k_m$ and $r_{12} \ldots . r_{1m}$ from equations (14) and (15) in equation (9) yields

$$\mathrm{R}^2 s_1{}^2=4(s_2{}^2 \sin \beta_2 \sin \gamma_2+ \ldots . s_m{}^2 \sin \beta_m \sin \gamma_m), \quad . \; . \quad (16)$$

and R can thus be determined directly from $2(m-1)$ measured angles of deflexion.

In the operation of the machine its physical limits should naturally be taken into account. From equation (15) we note that

$$\sin \gamma_2 = \frac{r_{12}s_1}{2\,s_2} = \frac{\Sigma xy}{2\Sigma y^2},$$

and since the sine of an angle cannot be greater than unity, it follows that Σxy should not be larger than $2\Sigma y^2$, and in case it is, it would be necessary to choose a larger unit for expressing the deviations $x_1 \ldots . x_n$, thus reducing the value of Σxy. As a matter of fact, in order to ensure the greatest degree of accuracy, there should always be a judicious choice of units so that the angles of deflexion are neither too large nor too small.

5. *The Labour of Mechanical Computation.*

Let us once more consider the case of nine independent variables, each series consisting of 25 items.

In order to determine the values of the k's it is necessary to move 9×25 pulleys along the arms of the nine rotating beams and adjust them to their proper positions. Also the 25 hooks must be moved and adjusted. The total number of mechanical adjustments to be made is, therefore, $25\,m$ or 250.

If the machine is properly constructed each adjustment consists of loosening a screw, shifting of a pulley or hook to its proper position along a graduated arm and tightening the screw. Each adjustment could be comfortably performed in the space of ten seconds, so that the total time required for the complete adjustment is $\frac{25\,m}{6}$ minutes.

The next operation is to find the values of $r_{12} \ldots . r_{1m}$ or $\gamma_2 \ldots . \gamma_m$, but this determination can be performed very rapidly, so that a liberal estimate of the time t_2 for making all the necessary adjustments, taking the readings and carrying out the final calculation of R would be roughly

$$t_2 = 6m \text{ minutes} \quad \ldots \ldots \ldots \quad (17)$$

If $m=10$, then $t_2=60$ minutes or one hour, which compares extremely favourably with the 750 minutes required for the numerical computation.

Below is a table showing, for different values of m, the value of t_1/t_2, which is the ratio of the times required for the numerical and mechanical methods of computation respectively. The value of this ratio according to equations (10) and (17) is

$$\frac{t_1}{t_2} = \frac{5}{6}m + \frac{1}{24}m^2, \quad \ldots \ldots \ldots \quad (18)$$

and although this equation cannot be regarded as strictly accurate, the table nevertheless gives a fair indication of the labour saved by the mechanical device.

TABLE.

m.	t_1/t_2.
3	2·87
4	4·00
5	5·20
6	6·50
7	7·87
8	9·33
9	10·87
10	12·50

6. *Theory Dealing with the Stability of the Rotating Beams.*

We have seen that when the instrument has been adjusted, it attains a state of equilibrium, each beam taking up a definite position. As we now wish to deal with the stability of the beams and as we are consequently obliged to consider variations in the angles of deflexion, we must distinguish between the positions of equilibrium of the beams in their undisturbed and disturbed positions. Let the undisturbed angles originally denoted by $\beta_2 \ldots . \beta_m$ now be denoted by ${}_0\beta_2 \ldots {}_0\beta_m$, and let the angles in any disturbed position be generally denoted by $\beta_2 \ldots . \beta_m$.

With the modified notation equation (13), for the case of m variables, may be rewritten :—

$$\left.\begin{array}{l} r_{12}s_1=2(s_2\sin {}_0\beta_2+s_3r_{23}\sin_0\beta_3+\ \ldots .\ s_mr_{2m}\sin_0\beta_m), \\ r_{13}s_1=2(s_2r_{23}\sin {}_0\beta_2+s_3\sin_0\beta_3+\ \ldots .\ s_mr_{3m}\sin_0\beta_m), \\ r_{14}s_1=\text{etc.} \end{array}\right\} \quad . \quad . \quad (13\,a)$$

Suppose we apply an extraneous moment by suspending a mass M from one of the beams at a distance L from its axis. Without loss of generality in the theory to be developed, we may choose B_2 as the beam to which the mass is attached. The moment of the extraneous force on B_2 is evidently equal to $-MgL\cos\beta_2$, and this added moment will disturb the original equilibrium of the system, causing a deflexion not only of the first, but also of all the other beams. The new equations of equilibrium of the beams now become

$$\left.\begin{array}{l} r_{12}s_1-\dfrac{ML}{2a\rho s_2}=2(s_2\sin\beta_2+s_3r_{23}\sin\beta_3+\ \ldots .\ s_mr_{2m}\sin\beta_m), \\ r_{13}s_1=2(s_2r_{23}\sin\beta_2+s_3\sin\beta_3+\ \ldots .\ s_mr_{3m}\sin\beta_m), \\ r_{14}s_1=\text{etc.} \end{array}\right\} . \quad (19)$$

If (19) is subtracted from (13 a) the result is

$$\left.\begin{aligned}
-\frac{\mathrm{ML}}{4a\rho s_2} &= s_2(\sin {}_0\beta_2-\sin\beta_2)+s_3r_{23}(\sin {}_0\beta_3-\sin\beta_3)+\dots\\
&\qquad\qquad +s_mr_{2m}(\sin {}_0\beta_m-\sin\beta_m),\\
0 &= s_2r_{23}(\sin {}_0\beta_2-\sin\beta_2)+\dots s_mr_{3m}(\sin {}_0\beta_m-\sin\beta_m),\\
0 &= \text{etc.}
\end{aligned}\right\} \quad . \quad . \quad (20)$$

We may now write down the solution for $s_2(\sin {}_0\beta_2-\sin\beta_2)$, obtained from this last set of equations, thus :

$$s_2(\sin {}_0\beta_2-\sin\beta_2)\Delta=\frac{\mathrm{ML}}{4a\rho s_2}\Delta', \quad . \quad . \quad . \quad (21)$$

where

$$\Delta=\begin{vmatrix}
1 & r_{23} & \dots\dots & r_{2m}\\
r_{23} & 1 & \dots\dots & r_{3m}\\
\cdot & \cdot & \cdot\ \cdot\ \cdot\ \cdot & \cdot\\
\cdot & \cdot & \cdot\ \cdot\ \cdot\ \cdot & \cdot\\
r_{2m} & r_{3m} & \dots\dots & 1
\end{vmatrix},$$

and

$$\Delta'=\begin{vmatrix}
1 & r_{34} & \dots\dots & r_{3m}\\
r_{34} & 1 & \dots\dots & r_{4m}\\
\cdot & \cdot & \cdot\ \cdot\ \cdot\ \cdot & \cdot\\
\cdot & \cdot & \cdot\ \cdot\ \cdot\ \cdot & \cdot\\
r_{3m} & r_{4m} & \dots\dots & 1
\end{vmatrix}.$$

But if we denote by r_2 the multiple correlation coefficient of the first independent variable on all the rest of the independents, it will be seen from (9 a) that

$$(1-r^2)\Delta'=\Delta,$$

which in combination with (21) yields

$$\mathrm{ML}=4(1-r_2{}^2)a\rho s_2{}^2(\sin {}_0\beta_2-\sin\beta_2). \quad . \quad . \quad . \quad . \quad (22)$$

If we choose M such that it is just sufficient to bring the beam B_2 to its horizontal position, and put $\mathrm{M}=\mathrm{M}_2$ when $\beta_2=0$, then

$$\mathrm{M}_2\mathrm{L}=4(1-r_2{}^2)a\rho s_2{}^2\sin {}_0\beta_2. \quad . \quad . \quad . \quad . \quad (22\,a)$$

Let us consider how the last equation is to be interpreted. If r_2 is nearly equal to unity, *i. e.*, if there is nearly complete correlation between the first independent and the rest of the independents, then a small applied moment will be sufficient to bring the beam to its normal position, which means that the beam B_2 is practically in neutral equilibrium, as one is inclined to expect. In general, equation (22) shows that the larger the value of r_2 the less stable is the beam concerned, and by finding the value of the mass M_2 required to bring the beam to its normal horizontal position one is able to calculate the value of r_2 directly.

Having established the relation between r_2 and the stability of the beam concerned, our next step will be to calculate the alteration in the value of R when one of the independents is discarded.

According to (9) and (14)

$$R^2 s_1 = 2(s_2 r_{12} \sin {}_0\beta_2 + s_3 r_{13} \sin {}_0\beta_3 + \dots\dots s_m r_{1m} \sin {}_0\beta_m) . \quad . \quad (23)$$

Substituting the values of $r_{12} \dots\dots r_{1m}$ from (19) and rearranging the terms we find that

$$R^2 s_1^2 = \frac{ML}{a\rho} \sin {}_0\beta_2 + 4 s_2 \sin \beta_2 (s_2 \sin {}_0\beta_2 + s_3 r_{23} \sin {}_0\beta_3 + \dots\dots s_m r_{2m} \sin {}_0\beta_m)$$
$$+ 4 s_3 \sin \beta_3 (s_2 r_{23} \sin {}_0\beta_2 + \sin {}_0\beta_3 + \dots\dots s_m r_{3m} \sin {}_0\beta_m) + \text{etc.}$$

Hence by virtue of (13 *a*)

$$R^2 s_1^2 = \frac{ML}{a\rho} \sin {}_0\beta_2 + 2 s_1 (s_2 r_{12} \sin \beta_2 + s_3 r_{13} \sin \beta_3 + \dots\dots s_m r_{1m} \sin \beta_m).$$

In the particular case when $\beta_2 = 0$ and $M = M_2$ the last expression in the brackets is, in analogy with (23), the expression for $\frac{1}{2} R^2_{-2} s_1$, where R_{-2} is the multiple correlation coefficient of the dependent on all the independents with the exception of the first. Hence

$$a\rho s_1^2 (R^2 - R^2_{-2}) = M_2 L \sin {}_0\beta_2 , \quad . \quad . \quad . \quad . \quad . \quad (24\,a)$$

and with the aid of (22) and (22 *a*) a more general expression may be derived,

$$a\rho s_1^2 (\sin {}_0\beta_2 - \sin \beta_2)(R^2 - R^2_{-2}) = ML \sin^2 {}_0\beta_2 . \quad . \quad . \quad . \quad (24)$$

The last equation may in certain cases be of considerable practical importance. It is often desirable to know what contribution the inclusion of any particular independent variable makes to the value of R. Thus in long-range weather forecasting one strives to obtain the largest possible correlation between the rainfall of a given area and the values, say, of previous pressures and temperatures over the same or different areas, the latter being treated as the independent variables. Suppose now one has determined the regression coefficients connecting the dependent variable with the independents, then it may happen that one or other of the latter does not contribute very materially to the value of R, in which case it may conveniently be discarded. In order to find out what difference the inclusion of any independent makes to the value of R, the obvious procedure would be to determine R with and without the said independent, but the result arrived at above provides a means of making a much more rapid calculation. In order to find the value of

$R^2-R^2_{-2}$ it is only necessary to suspend a suitable mass M from the beam B_2 at a convenient distance L from the axis, read off the new angle at which equilibrium is attained, and substitute the appropriate values in (23).

The theory dealing with the stability of the beams therefore provides us with a useful tool for practical correlation work.

Furthermore, elimination of M_2L between (22 *a*) and (23 *a*) leads to the expression

$$\begin{aligned} s_1{}^2(R^2-R^2_{-2}) &= 4s_2{}^2(1-r_2{}^2)\sin^2{}_0\beta_2 \\ &= s_2{}^2(1-r_2{}^2)k_2{}^2, \qquad \ldots \ldots \quad (25) \end{aligned}$$

a very useful result, as it shows the change in the value of the multiple correlation coefficient when any particular independent variable is discarded.

7. *Preliminary Experimental Results.*

In this paper it is by no means intended to describe the details of the construction of the apparatus or of the mechanical difficulties to be overcome. That can more appropriately be done in another publication.

However, in order to demonstrate that the theory elaborated above can also be translated into practice, a simple and rather crude machine was constructed, and a few of the experimental results will be given here. The results are such that it seems quite feasible to build a fairly elaborate machine, capable of dealing with a large number of variables, with the promise of complete success.

The preliminary experiments with the small model, consisting of two rotating beams each carrying only three arms, have shown that the main mechanical difficulty to be surmounted undoubtedly is the elimination of friction; and pulleys of the finest construction are essential if the machine is to attain a high degree of efficiency. The only other mechanical difficulty of importance is that presented by the extension of the strings due to changes of tension, but also this presents no insurmountable obstacle.

The results of three experiments are given below :—

	β_2.	β_3.	k_2.	k_3.	r_{12}.	r_{13}.	R.
1. Theoretical....	—	—	−1·126	−0·252	−0·962	0·127	1·00
Experimental..	−32½°	−7½°	−1·08	−0·26	−0·97	0·17	0·93
2. Theoretical....	—	—	−0·124	0·751	−0·50	0·994	1·00
Experimental..	−3½°	23½°	−0·12	0·80	−0·45	0·85	0·95
3. Theoretical....	—	—	0·687	0·880	0·240	0·795	1·00
Experimental..	20½°	27½°	0·70	−0·92	0·31	0·68	0·99

In view of the crudeness of the model and the fact that angles could only be read to the nearest degree, the agreement between the experimental and theoretical values may be regarded as quite satisfactory.

8. *The Solution of Simultaneous Linear Equations.*

By a slight adaptation, the apparatus may also be employed to solve any set of independent simultaneous linear equations. Suppose that m unknowns $u_1 \ldots . u_m$ have to be determined from the m equations

$$\left.\begin{array}{l} c_1 = a_{11}u_1 + a_{12}u_2 + \ldots . a_{1m}u_m \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ c_m = a_{m1}u_1 + a_{m2}u_2 + \ldots . a_{mm}u_m \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad (26)$$

The condition of independence of these equations is that the determinant

$$\Delta'' = \begin{vmatrix} a_{11} & \ldots . & a_{1m} \\ a_{21} & \ldots . & a_{2m} \\ \cdot & \cdot \quad \cdot & \cdot \\ \cdot & \cdot \quad \cdot & \cdot \\ a_{m1} & \ldots . & a_{mm} \end{vmatrix} \neq 0.$$

For the solution of m equations we require a machine with m rotating beams B_1 to B_m, each beam being provided with m arms.

In the first place, since the movable pulleys on the arms have to be adjusted at distances proportional to the constants, it would in general be necessary to multiply or divide each of the above equations by a suitable factor, so that the constants lie within the limits of the machine.

We now proceed exactly as before by adjusting the pulleys on the arms of B_1 at distances $a_{11}\ a_{21} \ldots . a_{m1}$ from the axis, on B_2 at distances $a_{12}\ a_{22} \ldots . a_{m2}$ etc. Rods F_1 to F_m are suspended from the ends of the strings the same as before, which causes the tensions in all the strings to be equal to, say T_0.

In the previous case we saw that the sum of the moments of the forces exerted by the strings on every beam, when in its "normal" position, was equal to zero. In this case, however, this condition no longer applies, because the constants in the equations are such that the algebraic sum of the lengths of the arms on any beam are no longer necessarily equal to zero. Hence, in order to obtain equilibrium it would be necessary to apply extraneous moments to each of the beams. This could be accomplished by suspending a mass W_1 from the first beam at a distance L from its axis, and choosing W_1 such that

$$W_1 g L = 2T_0(a_{11} + a_{21} + \ldots . + a_{m1}).$$

Similarly, a mass W_2 is attached to the second beam such that

$$W_2 gL = 2T_0(a_{12} + a_{22} + \dots a_{m2})$$

and so for the other beams.

All the beams will now be in equilibrium in their horizontal positions. Moreover, at whatever angle any particular beam B_f eventually comes to rest, the moment exerted by the mass W_f will always balance the sum of the moments of the normal tensions T_0 of the strings, and hence we need concern ourselves in the further discussions only with the *changes* in these tensions $T_1 \dots T_m$.

The procedure to obtain the values of $u_1 \dots u_m$ is now exactly as before. The hooks $E_1 \dots E_m$ are moved through vertical distances $c_1 \dots c_m$. This will cause the system to find a new state of equilibrium, the angles of rotation of $B_1 \dots B_m$ being $\beta_1 \dots \beta_m$ respectively.

Now consider the equilibrium of each beam in turn. In each case the moments of the forces must be equal to zero, and hence

$$\left.\begin{array}{l} a_{11}T_1 + a_{21}T_2 + \dots a_{m1}T_m = 0, \\ a_{12}T_1 + a_{22}T_2 + \dots a_{m2}T_m = 0, \\ \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\ a_{1m}T_1 + a_{2m}T_2 + \dots a_{mm}T_m = 0, \end{array}\right\} \quad \dots \quad (27)$$

but since the determinant $\Delta'' \neq 0$, it follows that

$$T_1 = T_2 = T_3 = \dots T_m = 0\,;$$

but

$$T_1 = a\rho g(2a_{11}\sin\beta_1 + 2\,a_{12}\sin\beta_2 + \dots 2a_{1m}\sin\beta_m - c_1),$$

hence

$$= c_1 2(a_{11}\sin\beta_1 + a_{12}\sin\beta_2 + \dots a_{1m}\sin\beta_m),$$

and, similarly,

$$c_2 = 2(a_{21}\sin\beta_1 + \dots 2_{2m}\sin\beta_m),$$

$$\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot$$

$$c_m = 2(a_{m1}\sin\beta_1 + \dots a_{mm}\sin\beta_m).$$

These equations are identical with (26), provided

$$\left.\begin{array}{l} u_1 = 2\sin\beta_1, \\ u_2 = 2\sin\beta_2, \\ \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\ \cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\ u_m = 2\sin\beta_m. \end{array}\right\} \quad \dots \quad (28)$$

The values of the unknowns are thus found to be double the sines of the angles of deflexion of the beams.

An interesting point to note is that in this case the tensions in the strings remain unchanged, and hence no mechanical difficulty due to stretching should be experienced.

In view of its ability to facilitate the solution of simultaneous equations, the usefulness of the machine is by no means confined to statistical work, but it should also find application in engineering and other branches of science.

9. *Other Machines of a Similar Type.*

A search of the literature subsequent to the development of the theory of the machine, which for the sake of convenience will be referred to as the " balance model," revealed the fact that there has apparently been no previous attempt at the direct mechanical computation of multiple correlation coefficients.

On the other hand, several mechanical appliances have been devised for the solution of sets of simultaneous equations. There is also an electrical machine designed by Mallock(1) and built by the Cambridge Instrument Company Limited, which, although it is primarily intended for the solution of simultaneous equations, can also be employed for the direct solution by the method of least squares of any set of equations of condition without first forming the normal equations. This, in effect, amounts to the same as finding the coefficients of a regression equation.

Mallock's calculating machine consists essentially of a set of A.C. transformers each with a number of windings, the number of turns in these windings being proportional to the coefficients of the linear equations. The machine may be operated from any alternating current mains, and with the proper arrangement of the coils, the fluxes in the transformers are proportional to the unknowns, and hence the unknowns can be determined directly by reading off these fluxes on suitable measuring instruments. Theoretically there is hardly any limit to the capabilities of the machine, but in actual practice such matters as the resistance in the windings and stray fluxes have to be compensated, and when the machine has to deal with a large number of coefficients its construction becomes very elaborate. In fact, both in its construction and in its operation it would appear that the balance model is much less elaborate than, and therefore preferable to, the electrical machine.

In a short note Vidal(2) describes a mechanism for the solution of simultaneous equations in which the unknowns are read off in terms of angles of rotation. The principle is somewhat similar to that of the balance model except that the pulleys and strings are replaced by systems of interacting racks and pinions.

A machine which is essentially similar to that of the balance model in the form in which the latter is employed for solving simultaneous equations is described by Wilbur (3), and a large model has been built at the Massachusetts Institute of Technology, Boston. Although the construction is fairly elaborate, according to Wilbur's report, it operates satisfactorily. However, the brief theory presented by the author does not appear to be adequate in all respects. Thus he obtains the equations from the constancy in length of the different steel tapes (which he uses in lieu of strings), but he devotes no attention to the equilibrium of the rotating slotted beams, a matter which should undoubtedly have been considered.

Apart from the inadequateness of the theory as presented, Wilbur's machine is designed for the solution of simultaneous equations only, and its applications are, therefore, much more limited than those of the new balance machine, and hence the latter should find a very wide field of usefulness.

In conclusion I wish to express my gratitude to Prof. J. T. Morrison of Stellenbosch for his helpful criticism of this paper.

References.

(1) R. R. M. Mallock, Proc. Roy. Soc. cxl. p. 457 (1933).
(2) P. C. Vidal, *Comptes Rendus*, ccii. p. 1748 (1936).
(3) J. B. Wilbur, Journal Franklin Inst. ccxxii. p. 1748 (1936).

ABSTRACT.

In order to avoid the tedium of the numerical computation of multiple correlation coefficients, a machine has been devised which provides a mechanical means of determining the constants of an equation of regression and hence of multiple correlation. Where a fairly large number of variables is involved the labour of numerical calculation is excessive, and an estimate of the relative times required for numerical and mechanical computation respectively shows that the machine should prove itself to be of great practical value.

Satisfactory experimental results were obtained with a preliminary rough model.

In a slightly modified form the machine may also be used for the solution of sets of simultaneous linear equations.

XXVII. *Note on the Distributions of Temperature and Vapour Pressure around a Horizontal Wet Cylinder.*

By R. W. Powell, Ph.D., Physics Department, The National Physical Laboratory, Teddington, Middlesex *.

[Received August 14, 1939.]

Abstract.

Experiments are described in which unventilated thermocouple psychrometers are used to determine the temperature and vapour pressure distributions above, below, and to one side of the wet surface of a horizontal cylinder. The presence of a convection current passing over the surface in a downwards direction is revealed when the surface is cooled to a temperature well below that of the air. When the surface temperature is raised, a temperature is reached at which very little movement of the air occurs. Under the conditions of the present experiment this occurred when the surface temperature was just over 1° C. below the temperature of the surroundings. With further increase in surface temperature the convection currents change to an upward direction. It is shown that the rate of evaporation falls to a minimum value in the region of little air-movement, and, on the assumption that the whole of the moisture loss takes place by diffusion, the experimental data for the vapour pressure gradients near the wet surface are seen to lead to values of the correct order of magnitude for the coefficient of diffusion of water vapour into air.

Introduction.

The temperature distribution established by natural convection in the neighbourhood of a heated horizontal cylinder has frequently been investigated $^{(1), (2)}$, but no such exploration appears to have been made for the case in which the surface of the cylinder is saturated with water. Some little time ago the present author was interested in the use of thermocouples for psychrometric measurements$^{(3)}$, and it was then apparent that the need for forced ventilation of the wet thermocouple decreases as the diameter of the thermocouple wire is diminished. Using thermocouples composed of nichrome and constantan wires of

---

* Communicated by the Author.

No. 40 S.W.G., it was found that the wet-bulb depression, in the absence of any forced ventilation, amounted to within 2 to 3 per cent. of the maximum depression for the fully ventilated condition. The results of an exploration of the temperature distribution around a horizontal wet thermocouple in " still " air were given in fig. 3 of that paper, and, in the course of a later investigation into the natural evaporation of water from a horizontal cylinder, it was subsequently considered that the inclusion of a similar set of measurements taken with a small diameter wet-bulb thermocouple would enable the distribution of vapour pressure around the wet cylinder to be determined to at least a first order of approximation. The results of these subsequent experiments are described in the present paper.

Description of Apparatus.

A section of brass tubing about 0·8 cm. in diameter and 31 cm. in length was coated internally and externally with shellac and baked. A uniformly spiralled, closely fitting heater of nichrome wire was then inserted in the tube, and the outer surface wrapped with blotting paper and a double layer of tape. The heater was connected up in series with a 12 volt. battery, a variable resistance, and a standardized resistance. Measurement of the fall in potential across the standard resistance enabled the current flowing in the heater to be determined. When wet the external diameter of the cylinder thus formed was 0·92 cm. Its surface temperature could be measured by means of four butt-welded nichrome-constantan thermocouples, composed of No. 40 gauge wires, which were stitched into the tapes. To enable the normal evaporation measurements to be made when required, the cylinder was suspended from one arm of a balance so as to hang with its axis horizontal and at the centre of an enclosure measuring approximately 21 × 21 × 60 inches and having a removable window in each of the longer sides. Whilst the temperature and humidity explorations were being made the cylinder was held in a fixed position by means of two metal supports which made contact near its ends.

A number of thermocouple psychrometers were made up from the No. 40 gauge nichrome and constantan wires. In each instrument constantan wires of sufficient length to be brought outside the enclosure to an ice-pot were butt-welded to each end of a 6 cm. length of nichrome wire. A long length of nichrome wire was also welded to the centre of the 6 cm. wire. The composite wire was then mounted so as to stretch tautly across the insulated tips of two steel rods, about 12 cm. apart, the nichrome portion being mid-way between the two rods. The nichrome-constantan junction to serve as the wet bulb was wrapped with very fine cotton, which extended 1 cm. or so each side of the junction. When the two constantan leads and the central nichrome lead were brought to an ice-pot and there joined to copper wires, it became possible to connect them in pairs to a potentiometer, and thus to obtain readings for either the wet- or dry-bulb temperature or for their difference, the wet-bulb depression.

Experiment to determine Temperature and Vapour Pressure Distribution around the Wet Cylinder.

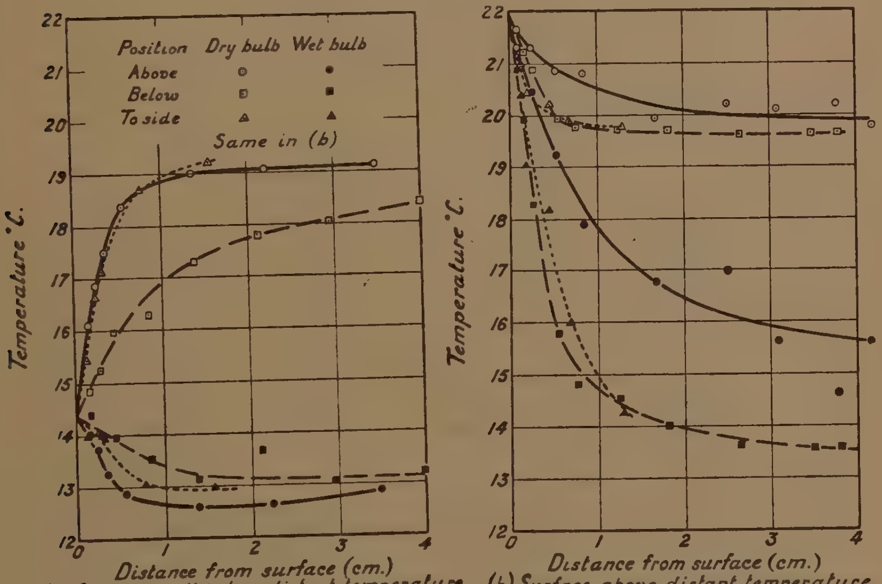
The first experiment to be carried out was an exploration to determine the wet- and dry-bulb temperatures in the air at various distances above, below, and to one side of the wet horizontal cylinder. Two of the psychrometers were mounted up in a horizontal position and attached to a supporting rod so as to have their junctions in the vertical plane containing the axis of the cylinder, but some 5·2 cm. apart. The supporting rod projected through the top of the enclosure, and raising or lowering this rod enabled one psychrometer to be used to record the conditions of the air immediately above the cylinder, while the other at the same time explored the conditions below the cylinder. It was also possible for the rod to be moved laterally, which allowed one of the psychrometers to be used for an independent investigation of the air conditions to one side of the cylinder.

In carrying out the test the surface and wet bulbs were well wetted with water, and readings commenced after a sufficient time had been allowed for equilibrium conditions to be established. Considerable fluctuations of the thermocouple readings were observed in certain positions, and in such cases the approximate mean value was noted. A correction of 2 per cent. was added to the wet-bulb depression to convert the wet-bulb temperature to a value corresponding to the fully ventilated condition.

The results obtained for the wet- and dry-bulb temperatures in two typical experiments are shown graphically in fig. 1. In experiment (*a*) no electrical heating was supplied, and the surface of the cylinder was cooled by evaporation to a temperature about 4·5° C. below its surroundings. In experiment (*b*) the wet surface was maintained some 2·4° C. above its surroundings by supplying electrical energy at the rate of 0·0524 watt per cm. length.

In condition (*a*) the rate of change of air temperature with distance from the surface is greater above the cylinder than below it, whereas the reverse is true for condition (*b*). These results are consistent with the presence of a downward movement of air in condition (*a*) and an upward movement in condition (*b*). A similar inversion is also noticed for the wet-bulb readings, the greater rates of change being above the cylinder in (*a*) and below the cylinder in (*b*). Smooth curves have been drawn through the various sets of points plotted in fig. 1, and the vapour pressures at various points have been calculated from wet- and dry-bulb temperatures read off from these curves. The distributions of vapour pressure in the air above, below, and to one side of the wet horizontal cylinder are shown in fig. 2 for conditions (*a*) and (*b*). It will be observed that the lines drawn through the experimental values for the vapour pressure converge to a common value at the surface equal to the saturation vapour pressure at the surface. This indicates the method used for the psychrometric measurements to be reasonably accurate, despite its drawback in introducing a further quantity of moisture. These results are represented in an alternative manner in fig. 3, where, from the somewhat limited data, an attempt has been made to draw the lines of constant vapour pressure for the two conditions. The shapes of these lines again indicate the existence of a downward convection current for condition (*a*) and an upward convection current for condition (*b*).

Fig. 1.

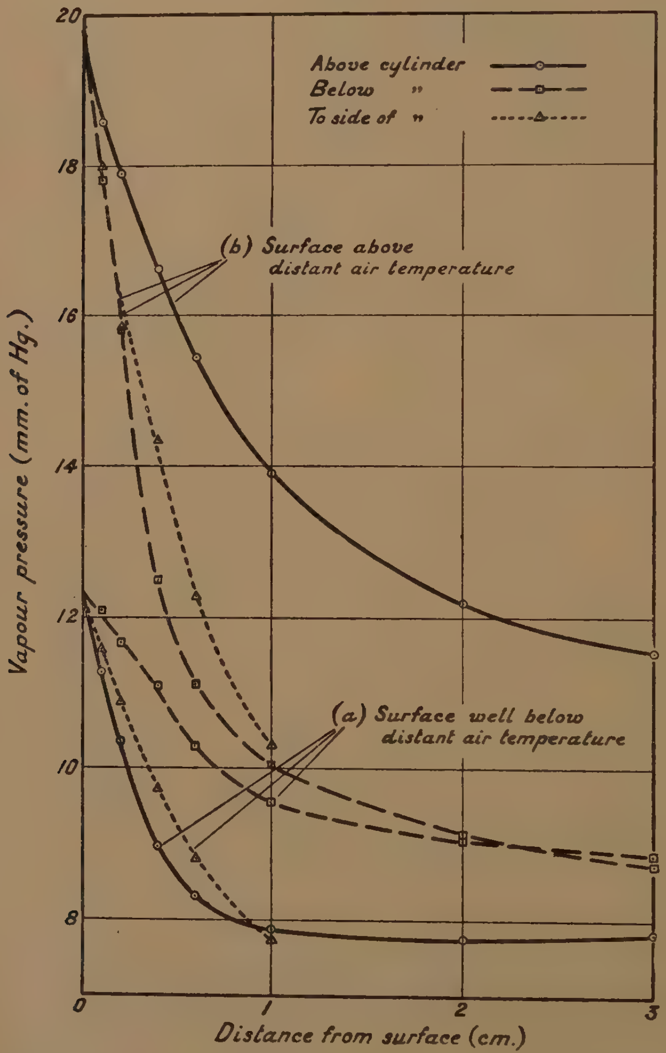


(*a*) Surface well below distant temperature (*b*) Surface above distant temperature.

Variation of wet- and dry-bulb temperatures of the air above, below, and to one side of a wet horizontal cylinder.

From the observed losses in weight during these two experiments a rate of evaporation of $7 \cdot 63 \times 10^{-6}$ gm. per sec. per cm. length of the cylinder was obtained for condition (*a*) and a rate of $18 \cdot 55 \times 10^{-6}$ for condition (*b*). The corresponding energy measurements show that for condition (*a*) the whole of the energy required to produce this rate of evaporation, namely 0·0185 watt/cm., must have been received by radiation from the surroundings and direct transfer from the air. For

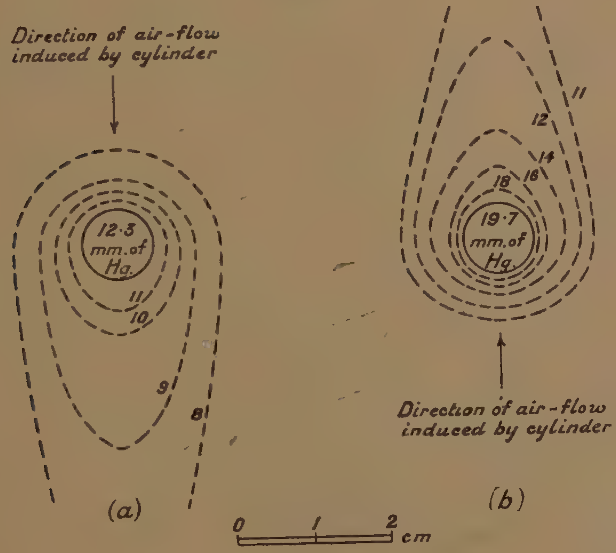
Fig. 2.



Variation of vapour pressure in the air above, below, and to one side of a horizontal cylinder with its surfaces (*a*) well below, (*b*) above distant air temperature.

condition (*b*) heat is lost from the surface by radiation and air transfer at the rate of 0·0074 watt/cm. length. A curve drawn through these values will be used subsequently in deriving approximate values for the rates of evaporation from measurements of the amounts of electrical energy supplied and the temperature difference between the wet surface and the surroundings.

Fig. 3.


Direction of air-flow induced by cylinder
12·3 mm. of Hg.
11
10
9
8
(a)
19·7 mm. of Hg.
18
16
14
12
11
Direction of air-flow induced by cylinder
(b)
0
1
2
cm

	(*a*)	(*b*)
Surface temperature	14·4° C.	21·9° C.
,, vapour pressure	12·3 mm. of Hg.	19·7 mm. of Hg.
Distant air temperature	$18{\cdot}9_5$° C.	$19{\cdot}5_5$° C.
,, ,, vapour pressure	7·8 mm. of Hg.	8·8 mm. of Hg.
,, ,, relative humidity	48 per cent.	52 per cent.

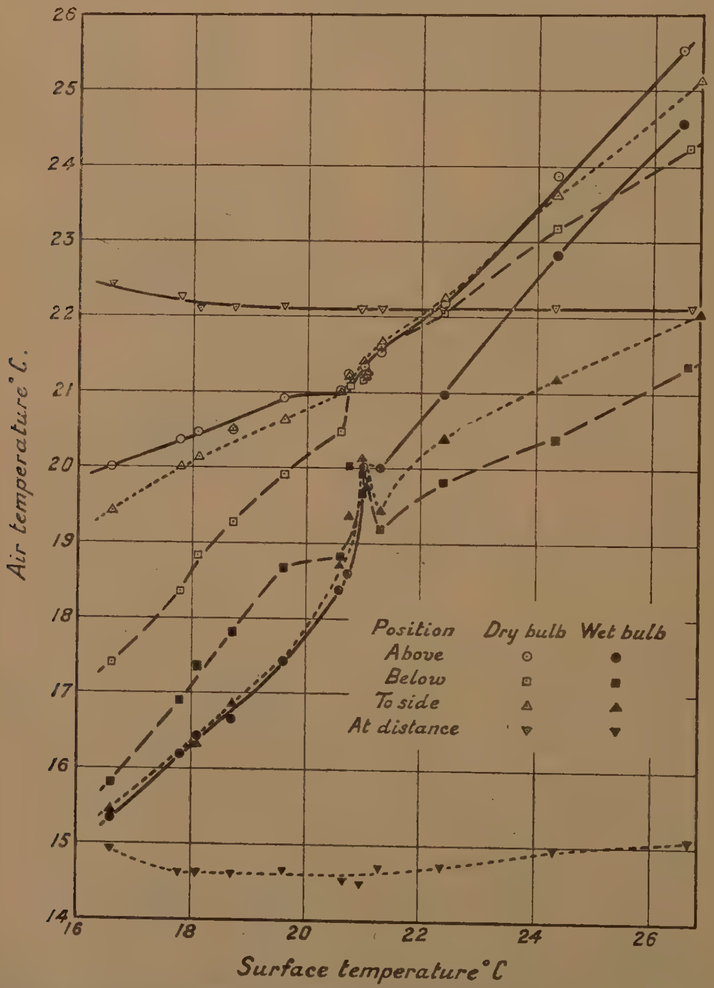
Approximate representation of lines of constant vapour pressure around horizontal wet cylinder for conditions (*a*) and (*b*).

More detailed Investigation of the Change in Direction of the Convection Currents.

The second experiment was undertaken with the object of investigating the temperature at which the change occurs in the direction of the convection current set up by the wet surface. The wet cylinder was firmly fixed in a horizontal position, and three thermocouple psychrometers mounted parallel to its axis and in positions 0·22 cm. above, below, and to one side of the surface of the cylinder. For these three positions readings similar to those of the first experiment were then obtained as the temperature of the wet surface was gradually increased.

The results obtained for the various wet- and dry-bulb temperatures are shown in fig. 4, and those for the corresponding vapour pressures in fig. 5. It is now seen that, for the particular air conditions of the experiment, the convection current inversion occurs when the wet surface is at a temperature of the order of 1° C. below that of its surroundings. The effect of this inversion was most apparent in the sudden change observed in the value of the wet-bulb depression of the psychrometer situated

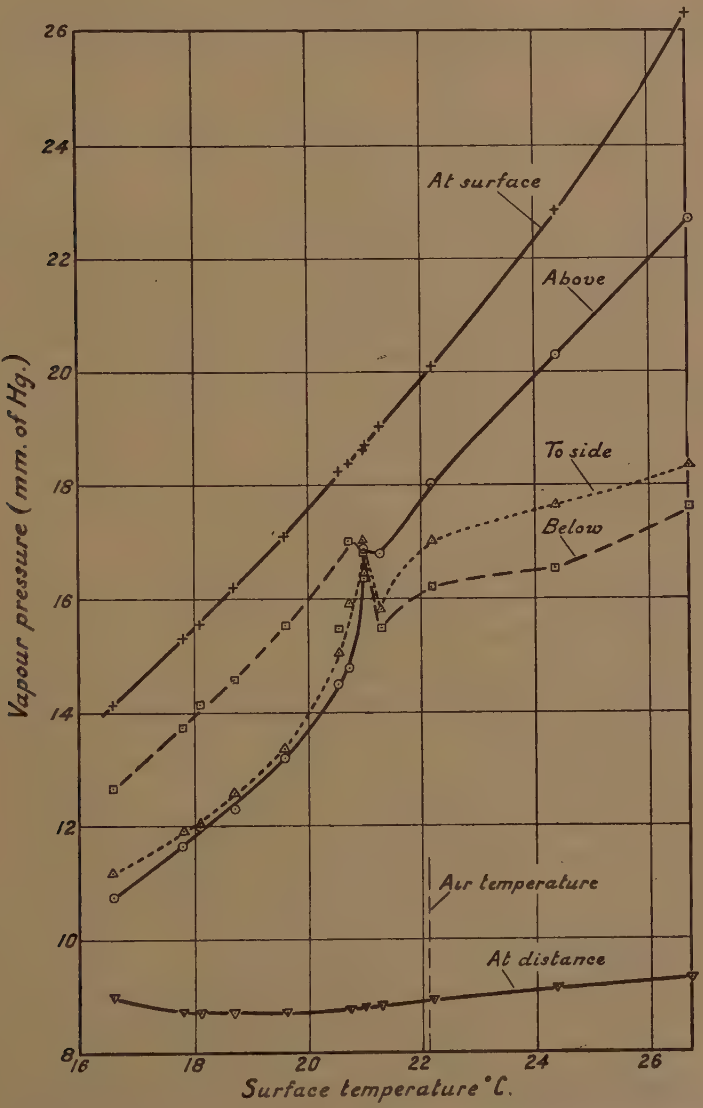
Fig. 4.



Wet- and dry-bulb temperatures 0·22 cm. above, below, and to side of wet horizontal cylinder at various surface temperatures.

above the cylinder. With the surface temperature at 20·75° C. the depression was 2·64° C., but the depression fell to 1·3° C. when the surface

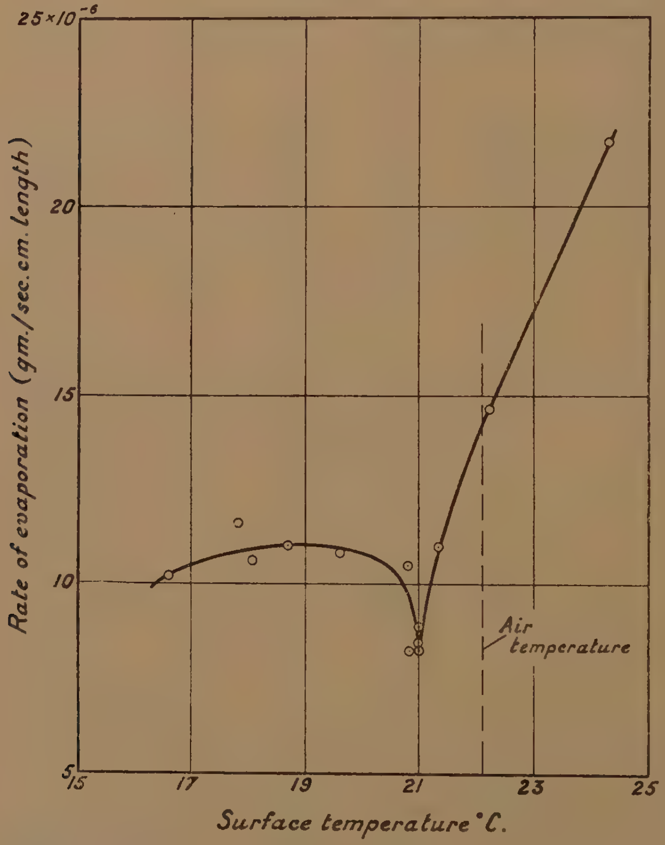
Fig. 5.



Vapour pressures 0·22 cm. above, below, and to side of horizontal wet cylinder at various surface temperatures.

reached 21·0° C. As the inversion region was approached, it was also noticed that the surface temperature of the cylinder tended to rise at an increasing rate, and in order to steady up at the temperature of the inversion it was necessary to reduce considerably the quantity of energy supplied to the cylinder. This is no doubt due to the rate of evaporation becoming smaller at this temperature owing to the cylinder becoming surrounded by relatively still air through which the water vapour has to pass to a greater distance by diffusion only. The nature of the change which occurs in the rate of natural evaporation as the surface temperature is increased is well shown in fig. 6. The actual points plotted have been obtained from the measured electrical energy supplied to the cylinder after applying corrections to allow for the transfer of additional heat by radiation and convection, and in some cases for the small variation of

Fig. 6.



Variation of rate of natural evaporation from horizontal wet cylinder with surface temperature.

temperature of the cylinder with time. These corrections are quite large, and the points plotted in fig. 6 should be regarded as approximate values. The curve is intended to show only the manner in which the rate of evaporation assumes a minimum value at the temperature for which the natural convection current changes its direction. A similar minimum was noticed by Hilpert⁽⁴⁾ in his experiments on evaporation from a vertical plate under still air conditions.

Estimation of the Coefficient of Diffusion of Water Vapour into Air.

On the assumption, that, under the conditions obtaining at the equilibrium state of minimum convection, the whole of the moisture transfer from the surface to the point at which the vapour pressure of the air has been measured takes place by pure diffusion, it should be possible to derive a value for D, the rate of diffusion of water vapour into air. At this equilibrium position the lines of constant vapour pressure approximate to circles, and for such radial flow D is given by the equation

$$D=\frac{Q\log_{10}\frac{r_2}{r_1}}{2{\cdot}73(C_1-C_2)}.$$

In this equation Q is the mass of water evaporated each second from a cm. length of the cylinder, r_1 is the radius of the cylinder, $r_2=r_1+d$, where d is the distance from the surface at which the measurements are made, and C_1 and C_2 are the corresponding concentrations of water vapour in the air. In the present instance Q is seen, from fig. 6, to have a value of $8{\cdot}2\times10^{-6}$ gm., r_1 and r_2 have values of 0·46 and 0·68 cm. respectively, and the data plotted in fig. 5 give a value of $1{\cdot}85\times10^{-6}$ gm. per cm.^{3} for (C_1-C_2). These quantities lead to a value of 0·28 cm.^{2} per sec., for the rate of diffusion. A repetition of the experiment under a different set of conditions gave a value of 0·20 cm.^{2} per sec.

It is not suggested that the present method offers a satisfactory means of determining the diffusion coefficient of water vapour into air. The accepted value for this coefficient is 0·25 cm.^{2} per sec. at 21° C.⁽⁵⁾, and it is satisfactory to find that the observations lead to values of the correct order of magnitude for this constant, despite the obvious limitations of the thermocouple psychrometer in that its presence must disturb the air conditions to a certain extent.

Discussion of Results.

Convection currents of the type described above are to be anticipated from a consideration of the local variations of density which arise from changes in the temperature and humidity of the air.

At normal atmospheric pressure, the density of the distant air, which has a temperature of 22·1° C. and vapour pressure of 8·7 mm. of mercury, is 0·001191 gm. per cm.^{3} The surface temperature which gives this same density for saturated air is 20·7° C. Lower surface temperatures correspond to a greater density and the air falls away from the surface and sets up a downward convection current. When the surface was at 20·7° C. the air 0·2 cm. to one side was found to have a temperature of 21·0° C. and a vapour pressure of 15·2 mm. of mercury, and therefore had a density of $0{\cdot}001191_5$ gm. per cm.^{3} It seems reasonable to suppose that it was this slightly higher density of the air at a small distance to the side of the cylinder which made it necessary for the surface of the cylinder to be heated to almost 21° C. before the equilibrium condition was reached. At surface temperatures above 21° C. the density of the air in the neighbourhood of the wet surface is definitely less than that of the distant air, and an upward convection current results.

The foregoing results have a bearing on problems of food storage and transport. In the absence of a forced air circulatory system, there will be a tendency for a concentration of moisture towards the ceiling of a chamber containing evaporating bodies with surfaces above, or even slightly below, air temperature, and a corresponding tendency for a concentration of moisture near the floor, should the evaporating surfaces be cooled considerably below the air temperature.

The results also demonstrate the difficulty of predicting the rate of evaporation from a surface in relatively still air, where the evaporation depends so much on the nature of the local air conditions, and where an increase in surface temperature, which normally leads to an increased rate of evaporation, may at times lead to a decreased rate.

Acknowledgments.

These experiments formed part of an investigation carried out at the National Physical Laboratory on behalf of the Engineering Committee of the Food Investigation Board. The author is indebted to the Director of the Food Investigation Board for permission to publish this account. He is also deeply grateful to Mr. M. J. Hickman, who constructed the apparatus, and rendered valuable assistance with most of the experimental observations.

References.

(1) B. Ray, Proc. Indian Assoc. Cultivation Sci. vi. p. 95 (1920).
(2) R. Hermann, ‘Forschungsheft,’ no. 379 (1936).
(3) R. W. Powell, Proc. Phys. Soc. xlviii. p. 406 (1936).
(4) R. Hilpert, ‘Forschungsheft,’ no. 355 (1932).
(5) ‘International Critical Tables,’ vol. v. p. 62.

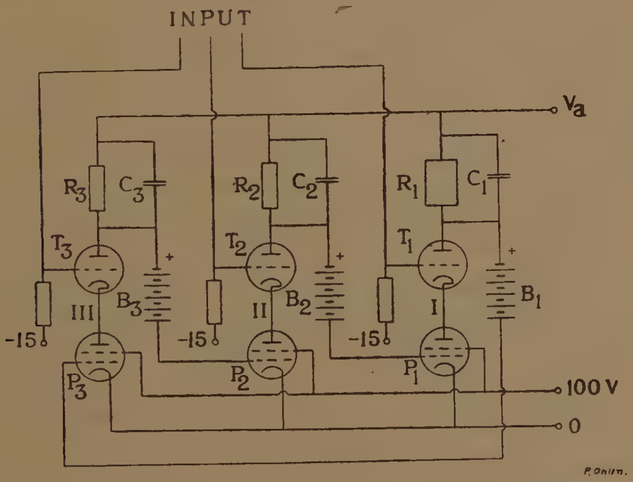
XXVIII. *A "Scale of Three" Counter.*

By P. Ohlin, Fil.Lic.*

[Received March 31, 1939.]

The most common method of counting rapid impulses consists of using "scales of two," and many such circuits of different construction have been published during the last years.

These counting circuits are characterized by having two stable conditions I and II. If the circuit is in the condition I an impulse will bring it to the condition II, and if it is in the condition II it will be brought to the condition I by the impulse.



Scales of two are good means of counting impulses, but it is often rather difficult to get them to work reliably. One of these difficulties is that the time it takes for the circuit to move from one condition to the other must be exactly adjusted to the duration of the impulses. In this respect the difficulty will be reduced by using a "scale of three," that is, a circuit having three stable conditions I, II, and III. Such a circuit works in the following manner: if, for instance, it is in the position I an impulse will bring it to the position II. The next impulse will move it to the position III, and a third one will bring it back to I, etc.

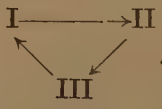
---

* Communicated by Prof. A. E. Lindh.

The figure shows such a "scale of three." It consists of three branches in parallel I, II, and III. Every branch is composed of a thyratron T and a penthod P in series. As shown in the figure, the grid bias of the penthod P_1 is determined by the plate-voltage V_a of the thyratrons, by the tension V_{B_2} of the battery B_2, which is the same for all three batteries B, and finally by the potential drop V_{R_2} over the resistance R_2. The grid bias V_{g_1} is thus$=V_a-B_{B_2}-V_{R_2}$. In this case $V_a=V_B=100$ V, and consequently $V_{g_1}=-V_{R_2}$, that is $V_{g_1}=0$ if no current flows through the branch II, as then $V_{R_2}=0$.

On the other hand, if a current flows through II V_{g_1} is negative, and by giving the current a sufficient value V_{g_1} gets negative enough to block P_1 completely. In the same way the grid bias V_{g_2} of P_2 is determined by the branch III and the grid bias V_{g_3} of P_3 by the branch I.

Suppose now that a current flows through the branch I. Owing to the potential drop over R_1 the grid of P_3 is negatively blocked and no current can pass through T_3 and P_3, that is T_3 cannot strike. The grid bias of P_2 is 0, then no current flows through III and there is no potential drop over R_3. The circuit is therefore in the following condition : the thyratron T_1 is lighted, the thyratron T_2 can be lighted by a positive impulse on its grid, but the thyratron T_3 cannot strike because of the blocking of the penthod P_3. If a positive impulse comes at the same time to the grids of the three thyratrons the following will happen : the branch I which is already lighted will not be influenced by the impulse, and the branch III, blocked by the negative grid bias of P_3, cannot strike. The branch II, however, where the grid bias of P_2 is about zero, will strike. Then the potential drop over R_2 will block the branch I, and T_1 will be extinguished. Hence the potential drop over R_1 will cease, and the grid bias of P_3 becomes 0. The impulse has therefore struck T_2, which was the only one that could be struck. The lighting of T_2 has caused extinction and blocking of T_1, and this extinction of T_1 has broken the blocking of P_3, and now T_3 is ready to strike. The next impulse will strike T_3, extinguish and block T_2, and make T_1 ready for striking. Apparently the circuit has three stable conditions, I, II, and III, characterized by the lighting of the thyratron T_1, T_2, T_3 respectively. Every impulse moves the circuit from one condition to the other according to the following scheme :



The greatest disadvantage of the thyratrons is the difficulty of extinguishing them. It can only be made by giving the plate voltage a sufficiently low value. Generally a thyratron can be lighted for plate voltages down to about 15 volts.

By the " scale of three " a thyratron will be extinguished by blocking the plate current. This blocking will last until the next thyratron has been lighted by a new impulse from the Geiger-Müller counter. Thus the duration of the blocking must always be at least as long as the duration of two impulses, and generally longer. Only if two impulses arrive immediately after each other the blocking becomes as short as the duration of two impulses. The fact that a thyratron which has just been extinguished is not immediately ready for striking will cause a safe working of the circuit.

By a " scale of two " of a similar construction, for example one consisting of two thyratrons in parallel, which alternately are struck and extinguished, the extinction must be caused by a short impulse generated in some way by the impulse from the Geiger-Müller counter. This impulse must be of about the same duration as the Geiger impulse itself, for the thyratron must be ready to light immediately again. It is here the difficulties of getting a safe working will appear. This may be applicable to most types of " scales of two," at least to those containing thyratrons.

The " scale of three," on the other hand, will make use of all the three conditions which are possible for a thyratron, that is :

(1) The thyratron is lighted, (2) it is blocked, (3) it is extinguished, but can be lighted. Every thyratron will in order pass these three conditions, and thus every risk is eliminated that a thyratron which has just been extinguished strikes again for the same impulse. By a " scale of two " which makes use only of the conditions (1) and (3) this trouble may appear.

To get the " scale of three " working safely one must observe the following conditions. Suppose that T_1 is lighted when an impulse arrives to the circuit. As we have mentioned above, this impulse will strike T_2 and at the same moment T_1 will be blocked, and this blocking will make T_3 ready for lighting. This will happen during a very short time, probably in a few micro-seconds. It may therefore occur that the impulse from the Geiger-Müller counter, having a duration of 10^{-3}–10^{-4} sec., will last until the blocking of T_3 has ceased. In that case T_3 will strike, that is, the circuit will count the impulse twice. Therefore it is necessary to delay the resetting of T_3. This will be accomplished by the capacities C in the following way :—

When T_1 becomes blocked the potential drop over R_1 will cease and the grid voltage of P_3 will rise to about 0. This will not happen instantly, however, because the condenser C_1 must be discharged through R_1. With a convenient value of $R_1 \times C_1$ the blocking of T_3 will last until the impulse has died, even if T_2 strikes instantly.

The thyratrons used above are Philips 4686 and the penthods are American valves, type 41. The resistances R are all$=3000\ \Omega$, and the capacity of C $1\ \mu$F, which gives $R\times C$ the value 3×10^{-3} sec., that is, the circuit can count about 350 imp./sec. It is possible, however, to increase the counting speed of the circuit several times by giving $R\times C$ a smaller value.

The Geiger counter impulses are amplified by a one-step resistance amplifier, which also makes them positive. The amplified impulses arrive at the grids of the thyratrons at the same time (by INPUT in the figure). To avoid capacitative coupling between the grids of the thyratrons three amplifier valves in parallel are used, one for each thyratron. The grids of these valves are connected directly to each other and by a capacity of 50 cm. to the Geiger counter. The mechanical counter is placed in series with T_1 (R_1 in the figure is the mechanical counter) and can count 125 imp./sec.

The current through the branches is determined by the screen grid bias of the penthods. The screen grid bias here used is 100 volts, which gives a plate current of 15 mA, sufficient to make the mechanical counter work. As R is $3000\ \Omega$, the potential drop over R is 45 volts, that is, the grid bias of a penthod in a blocked branch is -45 volts. The grid bias of the thyratrons is taken from a battery and is -15 volts, and the grid leaks are $2\times10^5\ \Omega$. All the voltages can be varied some per cent. without disturbing the work of the circuit.

The “ scale of three ” has been used for some measurements of X-ray intensity by the Geiger counter, and has proved to work in a perfectly reliable way.

Physics Laboratory,
University of Uppsala.
March 1939.

XXIX. *On the Proposition that "Shearing Stress alone is effective in producing Elastic Failure."*[*]

By James J. Guest.

[Received July 17, 1939.]

Notation p_1, p_3, p_2 . . . Principal Stresses in descending Order of Magnitude.

In a number of recent papers the respective authors have set forth as a definite scientific fact that shearing stress alone is effective in producing elastic failure. The proposition is of considerable importance as, if true, it would simplify the problem of combined stress greatly, and the designer, who is faced by the confliction of weight and strength—as particularly in aircraft—would no longer be confronted by about ten conflicting criteria to select from as the basis of his calculations: they would be reduced to two.

In "The Plastic Distortion of Metals"[†], G. I. Taylor, F.R.S. (Research Professor in the University of Cambridge), and H. Quinney, M.A., after defining the "principal shear stresses" to be half the differences of the principal stresses, proceed: "The importance of the principal shear stresses lies in the fact that it is found experimentally that the addition of a total hydrostatic pressure to any stress system does not produce any plastic strain. This depends only on the shear stresses."

Having complete reliance upon this proposition, these scientists confine their investigations to those two theories of elastic failure which are in accordance with it, the culmination of their paper being entitled "Comparison between von Mises' first hypothesis for the stress at which Distortion begins . . . and Mohr's hypothesis," the former being that the shear energy content of the material is the limiting factor, leading to the equation

$$(p_2-p_3)^2+(p_3-p_1)^2+(p_1-p_2)^2=\text{const.} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and the latter—according to Taylor and Quinney—being identical with the shear stress theory, which states that, in all isotropic materials, elastic failure occurs when the maximum shearing stress reaches a certain amount, the expression of this criterion being accordingly

$$p_1-p_2=\text{const.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

[*] Communicated by Jas. G. Docherty.

[†] Phil. Trans. Royal Society, A, vol. ccxxx. pp. 323–362. See pp. 348–349.

These scientists give as their conclusions that "in the case of copper and aluminium it (*i. e.*, their experimental research) provides a striking confirmation of the von Mises' hypothesis," and with regard to mild steel that "it appears that von Mises' hypothesis is nearer the truth than Mohr's."

There have been a number of researches upon mild steel in which close agreement with the shear stress theory was found, and, possibly recognising that all these could not have been seriously in error, Professor Taylor later devoted a paper, * "Faults in a Material which yields to Shear Stress while retaining its Volume Elasticity," to the reconciliation of these very differently based theories, and finally gives as his verdict that "It may therefore be true in general that the law of elastic breakdown is that of Mohr for materials for which the stress drops at the yield point and that of v. Mises for materials which do not possess this property."

In "The Elastic Limit of Metals exposed to Tri-axial Stress" †, G. Cook, D.Sc. (Professor in the University of Glasgow) starts by describing the experiment which he carried out by superposing a high hydrostatic pressure upon a soft iron wire coiled into a helical spring and just overloaded. He found that the permanent set in the spring was somewhat reduced. Cook then proceeds thus, "the absence of any appreciable effect justifies the conclusion that the addition of such pressure does not increase the liability of this material to elastic breakdown" . . . "if the conclusion is accepted that the position of the elastic limit is independent of the fluid stress, it follows that stress-differences only will be found in the expression of any theory which can be applicable. The shear theory of Guest and the shear strain energy theory of von Mises-Hencky comply with this condition."

Later, in a paper entitled "The Effect of Fluid Pressure on the Permanent Deformation of Metals by Shear" ‡ Professor Cook describes a difficult, but more informative, variation of the above experiment, in which he looked for the effect of a superposed hydrostatic pressure on the actual yield point of the soft iron helical spring. He found that the effect was to reduce the load necessary to cause yield by a small amount (rather less than $1\frac{1}{2}$ per cent.), not in agreement with the previous result, and finally decided that "as regards the yield point the experiments appear to point quite definitely to the absence of any effect due to fluid pressure."

As strengthening the position, Prof. Cook writes § that "It is, of course, inconceivable that plastic flow would occur in isotropic material under simple fluid stress either tensile or compressive."

---

* Proc. Roy. Soc. vol. cxlv. p. 1 (Mar. 27, 1938). See p. 2.

† Proc. Roy. Soc. vol. cxxxvii. p. 559 (April 22, 1932).

‡ Inst. Civil Engineers—Selected Paper no. 170—Session 1934.

§ Proc. Roy. Soc. vol. cxxxvii. p. 561 (April 22, 1932).

The research upon Tri-axial Stress consisted in subjecting "thick" cylinders to combinations of material fluid pressure and axial stress, the thicknesses being such that the radial stress was of such intensity as to be comparable with the other stresses. Two steels, of different carbon content, were experimented upon, and Cook expresses his conclusions thus: "The evidence therefore favours the view that, as regards the initial break-down in mild steel, the maximum shear stress is the critical factor, that the Guest hypothesis is true to the extent that, for the same kind of stress distribution, the values of the principal stresses at the elastic limit are such as to give a constant value to the maximum shear stress."

Although these papers agree as to the proposition under discussion they disagree as to which of the criteria which conform with it is the true one.

In "Theories of Strength" *, A. Nadai (formerly Professor in the University of Göttingen) treats the proposition here considered as quite indisputable and states that evidence of the effect of volumetric stress was particularly looked for in the research, initiated by himself and continued by A. Lode, at the University of Göttingen, and was found to be absent. He states that the experimental results were, within 3 per cent., in agreement with the Mises-Hencky theory and that this is therefore to be accepted as the truth.

The research of Taylor and Quinney was carried out by subjecting tubes to various combinations of axial load and torque, and they present their results by plotting the observation points upon axes representing the axial stress and shear stress due to the applied axial forces and torques respectively, and drawing the curves corresponding to the two theories they consider upon the diagrams. Nadai criticizes their diagram for mild steel, pointing out that the observation points are far from being in satisfactory agreement with the ellipse representing the Mises-Hencky theory, but accept them as proof of its truth "because they lie nearer to this ellipse than to any other."

A formal presentation of the proposition is set out by Dr. H. S. Gough, F.R.S. (then Head of the Engineering Department of the National Physical Laboratory, now Director of Scientific Research at the War Office), and W. I. Clenshaw, B.Sc., in their paper upon "The Testing of Materials" †. This reads as follows: "It is now established that yielding

---

* Trans. Amer. Soc. Mech. Eng. vol. clv. p. 111 (1933).

† Proc. Inst. Marine Engineers, vol. xlvii. pt. x. p. 241 (1935). See pp. 261 & 262.

of ductile materials is not produced by hydraulic compression. Therefore, as the changing of a stress system denoted by p_1, p_2, p_3 to (p_1+x), (p_2+x), (p_3+x) produces no effect, it follows that certain criteria of yielding can be excluded from consideration; such criteria include the theory of Maximum Stress, Maximum Strain, and Constant Total Energy of Deformation. Hence the true criterion of yielding must involve principal stress differences, and suitable existing theories are reduced to three in number, namely, (*a*) Maximum Shear Stress Theory (usually termed Guest's Law in this country) . . . ; (*b*) Mohr's Theory * . . . if $p_1>p_3>p_2$ failure occurs at a value of $\frac{1}{2}(p_1-p_2)$ which must be a function of (p_1+p_2) whose form depends on the material in question; (*c*) Mises-Hencky Theory . . ." After discussing these three theories they write that "It may safely be concluded that the Mises-Hencky relation affords the most satisfactory criterion for the condition of yielding of ductile materials under combined stresses."

It is to be observed that, whatever the status of the "Mohr Theory" may be, the entrance of (p_1+p_2)—which cannot be expressed in terms of principal stress differences only—into the condition for yield, implies that some form of volumetric stress is involved. Hence Gough and Clenshaw are in error when they include it above.

The above citations display the important influence the proposition has had in guiding recent research and on the conclusions which the several authors have drawn from the results of their experimental work.

The results obtained in my research upon Combined Stress, however, indicated that volumetric stress of some type was an essential factor in the production of elastic failure, and this was included in the statement of my conclusions. The proposition under discussion is diametrically opposed to them, and, after making my position clear, I propose to discuss the mode whereby the above authors arrive at the conclusion that the proposition is true.

When Gough and Clenshaw make the implication that the Guest Law is identical with the Maximum Shear Stress Theory they misrepresent the position. In my paper on "Ductile Materials under Combined Stress" † the conclusions at which I arrived were expressed as follows:—"Any state of stress can be represented by the co-existence of a shearing stress, a volumetric uniform stress, and a stress perpendicular to the plane of the shear; and the yield point occurs when the shearing force together with a small fraction of the volumetric stress reaches a certain amount: the volumetric stress being reckoned positive when tensile."

---

* An exposition of the "Mohr Theory" is given by the present writer in the 'Automobile Engineer,' Feb. 1938.

† Phil. Mag. pp. 69–132 (July 1900).

"That the coefficient of the volumetric stress for steel is small has been shown by these tests, and that it has some value greater than zero seems probable, as volumetric tension would ultimately reduce the material to the liquid or gaseous form and cause the necessary shear to evanesce."

The equation of the criterion is given in the paper as

$$p_1-p_2+\lambda(p_1+p_2)=\mathrm{C}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where it will be observed that the volumetric stress employed is $\frac{1}{2}(p_1+p_2)$—and not the average volumetric stress, $\frac{1}{3}(p_1+p_2+p_3)$—and that this criterion differs from all the others discussed above in that it contains two constants of the material, C and λ, instead of one only. Of these C is a stress constant, being equal to twice the yield stress under pure shear.

Although the research was conducted on ductile materials only—steel, copper, and brass,—the variations in the ductility encountered were such as to enable a general law of failure applicable to all isotropic materials to be arrived at, and in its verbal expression accordingly all reference to ductility or brittleness is excluded. This property of materials is dealt with in the equation by the non-dimensional constant λ, which varies from zero to unity as the ductility gradually changes to brittleness, the view taken being that there is no sudden change or definite line of separation between brittle and ductile materials.

Being fully aware that engineers are disinclined to use any but the very simplest formulæ and bases of design, I included in my paper an approximate rule, for ductile materials only, to supplant the two theories then current, the Rankine or Maximum stress, and the St. Venant or Maximum elongation criteria, which were first shown to be definitely at variance with the facts by this research. The rule was worded thus :—" The result, then, of these experiments, as applicable to practice, is that the condition for initial yielding of a uniform ductile material is the existence of a specific shearing stress ; and that the intermediate principal stress is without effect."

Considering first the complete form of the " Guest Law," it is perfectly clear that, so far from being, as Gough and Clenshaw attempt to imply, identical with the Shear Stress Theory, it is in direct opposition to it, being a statement that the Shear Stress Theory is untrue and inadequate to account for the facts, and that the effect of volumetric stress of some type must be taken into account.

Neither is the simple rule—the Guest " hypothesis " as regarded by Professor Cook—in agreement with the Shear Stress Theory. The latter is advanced as a general law applying to all materials—hard stone, brick, and cast iron are specifically included*,—and as accurate for all.

---

* Phil. Trans. Roy. Soc. vol. clxxiii.

The simple form of the Guest Law, on the other hand, is specifically limited to ductile materials and is for practical design use only, as it is merely an approximation. In fact, to-day the progress of engineering requires the use of the complete law for fine design.

In any scientific discussion the complete form of the Guest Law alone may be considered : any suggestion that the simple rule was the complete criterion would thus be misleading. In a discussion on design procedure it would be otherwise, the matter then of importance being the amount of error involved in the use of the simple form of the criterion.

It is impossible that the view expressed above and the proposition here considered can both be true : the arguments advanced as proving the truth of the latter can now be discussed.

The fundamental statement is that "it is found experimentally that the addition of a total hydrostatic pressure to any stress system does not produce any plastic strain." While I hold that this cannot be universally true, it may be regarded as experimentally proven for such conditions as are of importance in the matter under discussion. Cook's first spring experiment is evidence of this nature and is typical of that afforded by other research.

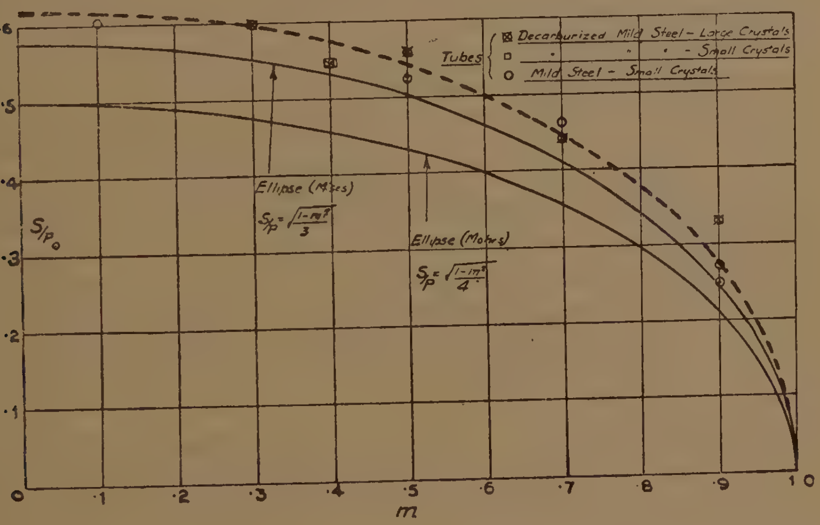
To argue hence that the superposition of a uniform volumetric stress is without effect is, however, illogical. If the effect of such a stress were to increase the shear stress necessary to cause yield, no effect would be apparent. This is what the Guest Law predicates. Even if it were a mere hypothesis, without the research from which it was deduced, it would render the argument untenable. However, as an illogical argument does not invalidate a deduction, the further evidence presented is to be considered.

Professor Cook's second spring experiment is an attempt to furnish such evidence, and the lowering of the yield point found, nearly $1\frac{1}{2}$ per cent., is the only research known to me in which a superposed hydrostatic pressure has been found to lower the resistance of the material in such tests. There are, however, many experimental results indicating the opposite effect. Thus P. W. Bridgman (Professor in Harvard University) submitted round steel bars, passing completely through a pressure vessel, to hydrostatic pressures so high as to cause the bars to fracture by "pinching off." The stress system produced in the material is there equivalent to a simple tension combined with a superposed hydrostatic pressure. The shear stress produced in the bars at failure exceeded that necessary in a simple tension test by 25 to 50 per cent.* Bridgman regards his results as completely disproving the shear stress theory. These very decided results suggest that there may be some explanation of the small contrary effect found by Professor Cook.

* Phil. Mag. (Jan and July 1912).

When it is stated that the application of a hydrostatic pressure to the outer surface of a body produces equal increments to the principal stresses already existing in it—as expressed in the presentation by Gough and Clenshaw,—the assumption is made that the body is isotropic and contains no cavities or inclusions. If the soft iron of the helical spring contained defects or inclusions, as it certainly would were it wrought iron (as is implied), the application of the hydrostatic pressure would, increasing the shear stresses at the discontinuities, tend to cause an earlier yield. I am not prepared to accept the result of this experiment in view of so simple an explanation.

The dictum of Professor Cook that it is inconceivable that volumetric stress, either tensile or compressive, would cause yield, is directly opposed to the view taken in my paper. The argument there that an increasing volumetric tension must eventually cause failure—at a stress $1/\lambda$ times the yield stress in pure shear according to equation (3)—is strongly against Cook's statement.



Proceeding now to the results for steel obtained experimentally by Taylor and Quinney, I consider that neither the author's conclusions that they support the von Mises theory, nor Nadai's argument to the same effect, are justifiable. In the figure the ellipse corresponding to the Beltrami-Haigh total energy theory, for a value of Poisson's ratio of 0·3, is drawn in broken line upon Taylor and Quinney's diagram * for steel. This presents the correlation between their experimental results and the

---

* As p. 350, fig. 12 *c*.

total and shear energy theories clearly. The Beltrami-Haigh ellipse lies fairly between the observation points, whereas the Mises ellipse lies entirely on one side of them, and for the most part at a distance from them exceeding what is likely to be the experimental error. Thus the results obtained by Taylor and Quinney for steel indicate that, if the criterion is to be an energy criterion, the energy of the volumetric stress is to be included. Their results are thus definitely against the proposition concerning the shear stress and against their conclusions, which are biased by their reliance upon it.

Neither do the results obtained by Professor Cook in his research upon tri-axial stress justify his conclusion that at yield the maximum shear stress is constant, for he qualifies his statement by the vague wording " for the same kind of stress distribution." This implies that the simple shear stress theory is inadequate to explain his results and that it is therefore not accurately true.

Examination shows that all the results obtained by Taylor and Quinney, and by Cook are in accordance with the Guest Law within the experimental error likely.

Professor Nadai, a strong advocate of the Mises-Hencky Theory, criticizes Cook's work as containing serious errors and makes a number of suggestions as to what these are. None of these suggestions is in any way sustainable, and Nadai omits to notice that several other carefully conducted researches, *e. g.* those of Professor Hancock [*], Middleton Smith [†], and Mason [‡] yielded results closely akin to those of Professor Cook, nor does he observe that the values of the ratio, ξ, of the yield stress in pure shear to that in simple tension determined experimentally have a wide variation—and not the one ratio corresponding to the Mises theory.

When Nadai writes that the search at Göttingen for evidence of any effect of volumetric stress revealed no trace of it, he is regarding the matter from a special point of view—that of the Mises-Hencky theory. Thus, if the value of ξ is found to be $\frac{1}{\sqrt{3}}$ $(=0{\cdot}577)$ it is taken as indicating that there was no volumetric stress effect. The same result, however, from the Guest criterion standpoint would be regarded as indicating a volumetric stress effect, measured by the value of λ being $0{\cdot}155$. Thus Nadai's statement is of doubtful import; it appears to be merely a different wording of the statement that the Mises-Hencky Theory was found to be correct within 3 per cent. From the energy point of view, the results for steel obtained by Taylor and Quinney would be regarded as showing that the full volumetric strain energy should be included, but Cook's results would be awkward so to account for they would imply a large negative effect of volumetric strain energy.

---

[*] Phil. Mag. 1905–1906.

[†] Proc. Inst. Mech. Eng (1909).

[‡] Proc. Inst. Mech. Eng. (1909).

These diverse results are strongly against the possibility of any criterion dependent on shear stress only as being true.

Furthermore, any criterion so dependent will necessarily predicate that all materials must have equal yield point stresses in simple tension and in simple compression as the shear stresses in these are identical. All tests of brittle materials, cast iron, stone, concrete, show yield stresses in compression several times those in tension, and it is only in mild and nickel steels, and soft metals, that experiment finds them to be nearly the same.

Thus any theory of elastic failure dependent on shear stresses only must be confined to such materials as have equal yield stresses in tension and compression and cannot be of general application.

As these scientists who have recently conducted research upon the subject have all independently stated, in the publications of important Institutions, that shear stresses alone are effective in producing elastic failure, it is probable that the proposition will be accepted as an established scientific fact by physicists and engineers generally. Against this acceptance I protest, deeming the proposition to be false and as leading to further erroneous and dangerous conclusions, and present the arguments given above in support of my contention.

XXX. *On the Interrelation of Molecular Constants for Diatomic Molecules.*—II.

By Dr. R. A. Newing,
University College of North Wales, Bangor *.

[Received August 14, 1939.]

1. In a previous paper ⁽¹⁾ attention was drawn to the need for a theoretical basis for the empirical relations between the nuclear distance r_e and the vibration frequency ω_e. Subsequent investigations ⁽²⁾ have shown that it is very difficult to produce specific results on account of the generality of the problem, and that the best hope of progress would be to relate the empirical results to some general physical principle. The general conclusion of these investigations has been that the empirical laws may be related to the existence, in the neighbourhood of the equilibrium configuration, of a common repulsive field for all similar molecules, provided that the laws are of the form

$$\mu\omega_e^2 f(r_e)=C, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

C being a constant for similar molecules. An unsuccessful attempt was made ⁽³⁾ to find a theoretical basis for a law not involving the reduced mass μ, and it was concluded that the methods employed had failed to produce evidence for a relationship of the form $\omega_e F(r_e)=C$. This latter conclusion is in agreement with a statement in a recent paper by Sutherland ⁽⁴⁾, although in my opinion the most that can be said at present is that the presence of the mass factor *appears* to be insuperable. Sutherland shows in his paper that the assumption of a mutual potential energy function

$$U=\frac{\alpha}{r^m}-\frac{\beta}{r^n}$$

leads to results similar in form to those of Badger ⁽⁵⁾ and of Allen and Longair ⁽⁶⁾, *i. e.*, of the form (1). It is not permissible, however, to regard this as providing a satisfactory basis for the empirical laws, for the constant C in Sutherland's work depends upon α or β, and, in order to provide a complete theoretical investigation, it would be necessary to show that one or other of these two quantities is approximately constant for similar molecules.

---

* Communicated by the Author.

[This remark does not apply to Sutherland's important result $D = kr_e^2/mn$, which does not depend upon α or β. It is to be noted that the derivation of a more accurate relation of this type would require a potential function which provided a good approximation, both in the neighbourhood of equilibrium and at large nuclear separations. In Sutherland's formula, D is not the observed dissociation energy, but is the value for the hypothetical vibrationless condition. The difference between these two quantities, however, is of the order of 1 per cent. in the molecular states considered by this author. The bond force constant k used in my previous paper (1) is one half the quantity as usually defined. Sutherland has used my definition in the derivation of his formula. With the accepted definition, *e. g.* as given by Badger (5), the relation becomes $D = kr_e^2/2mn$.]

2. In order to provide a physical basis for the empirical laws, use was made (2) of Slater's application of the Virial Theorem to molecular problems (7). He deduced the result that, if the zero-point vibration may be neglected,

$$\left.\begin{aligned} T &= -U - r\frac{dU}{dr}, \\ V &= 2U + r\frac{dU}{dr}, \end{aligned}\right\} \quad \dots\dots\dots \quad (2)$$

where U is the mutual potential energy of the nuclei fixed at a distance r apart, and T and V are the *mean* values of the electronic, kinetic, and potential energies, respectively, for this separation. The vibration frequency, or the bond force constant, will be known if $(d^2U/dr^2)_{r=r_e}$ is known. It follows from the above equation for T that

$$4\pi^2\mu\omega_e^2 = -\frac{1}{r_e}\left(\frac{dT}{dr}\right)_{r=r_e}, \quad \dots\dots \quad (3)$$

and, since T is a function of r, this will provide an equation of the form (1) where C and the function f may depend upon the particular molecular state considered.

On applying equations (2) to a standard type of energy function, Slater found that, in the neighbourhood of equilibrium, the decrease of energy responsible for molecular binding came from a decrease in kinetic energy rather than from a decrease in potential energy. In the neighbourhood of equilibrium the slope of the electronic kinetic energy curve is negative, indicating *repulsion*; that of the potential energy curve, $V(r)$, is positive, indicating attraction. In the above considerations, therefore, it follows that dT/dr is to be interpreted as the repulsive force, and we are led to propose the following general interpretation of the empirical laws represented by equation (1):

The empirical laws would arise from the existence of a common repulsive field for all similar molecules in the neighbourhood of the equilibrium configuration.

By investigating the properties of a simplified molecular model in which the electrons are considered as an electron gas, and in which the volume occupied by the valency electrons is proportional to r^n, it may be shown (2) that $d\mathrm{T}/dr$ takes on the form $ar^{-(2n+3)/3}$, where a is independent of r, but may depend upon the number of valency electrons and also upon the nature of the nuclear cores. The assumption of a common repulsive field (i. e., a constant for similar molecules) would lead to a law of the same form as that due to Allen and Longair ; or, conversely, the Allen-Longair rule would imply the existence of the common field.

Further empirical evidence for the existence of an approximately common repulsive field for similar molecules has been provided by the work of Huggins (8). He modified Morse's empirical potential function $\mathrm{U}(r)$ so that the repulsive term was the same for all electronic states of all particular molecules, and on choosing his constants to satisfy data from band-spectra, he was led to the conclusion that the repulsive field is nearly the same for similar molecules. The repulsive term in Huggins's potential function may not, of course, be identified with the kinetic energy, for his potential function is an approximation valid for all nuclear separations. Huggins does find, however, that his potential function requires modification for nuclear distances which are large in comparison with the equilibrium separation. This is understandable from our present point of view, for the common repulsive field is probably restricted to the neighbourhood of equilibrium configurations. Huggins showed that his results are in approximate agreement with the Badger law, and the present writer has also shown (2) that an approximate law of the form of (1) may be deduced from Huggins's potential function.

3. There has been no indication in this paper of the actual nature of the function $f(r)$, and the existence of an approximately common repulsive field has not been established theoretically, but our present investigations have led to the important result that there is a connexion between the laws (1) and the repulsive field. The former may be regarded as consequences of a common repulsive field, the existence of which is still to be established theoretically. Alternatively, the laws (1) may be regarded as empirical results giving information about the repulsive field ; further independent empirical evidence of the existence of such a common field is provided by the work of Huggins.

The advantage, in any general investigation, of using equation (3), which has been deduced from Slater's results, is that it is only necessary to consider the repulsive term in the neighbourhood of equilibrium configurations.

References.

(1) R. A. Newing, Phil. Mag. xix. p. 759 (1935).
(2) R. A. Newing, Ph.D. Thesis, University of London (1938).
(3) *Ibid.* pp. 107–109.
(4) G. B. B. M. Sutherland, Proc. Ind. Acad. Sci. A, viii. p. 341 (1938).
(5) R. M. Badger, J. Chem. Phys. ii. p. 128 (1934) ; iii. p. 710 (1935).
(6) H. S. Allen and A. K. Longair, Phil. Mag. xix. p. 1032 (1935).
(7) J. C. Slater, J. Chem. Phys. i. p. 678 (1933).
(8) M. L. Huggins, *ibid.* iii. p. 473 (1935) ; iv. p. 308 (1936).

XXXI. *The Effect of a Magnetic Force on High Frequency Discharges in Pure Gases.*

By A. E. Brown, B.A., B.Sc.,
Merton College, Oxford *.

[Received January 10, 1940.]

1. In a recent paper † on a generalization of the theory of electrical discharges, experiments have been described by Professor Townsend and Mr. Gill on the effect of a magnetic force on the high-frequency force necessary to start a discharge in a spherical bulb containing dry air. In this paper I propose to give an account of similar experiments which I have made with nitrogen and helium, having taken special care to purify the gases. The experiments with nitrogen were made in collaboration with Mr. G. O. Jones.

The nitrogen was prepared by heating sodium azide previously recrystallized in a sulphuric acid desiccator. The vessel used for storing the nitrogen was thoroughly cleaned and baked. It contained phosphorus pentoxide, which had been distilled into the vessel in a stream of ozonized oxygen, in order to eliminate impurities from the lower oxides of phosphorus. The vessel containing the gas had a side tube with two taps, so that the small amount of gas in the space between the taps could be let into the bulb in which the discharge took place. This bulb (B, fig. 1) was of quartz, of about 13 cm. diameter, and was evacuated and heated to 600° C. for several hours to remove all adsorbed impurities. The bulb had two quartz side tubes in the form of traps which could be surrounded with liquid air. One was empty, and when cooled served to prevent any volatile constituents of the tap grease reaching the discharge bulb. The other contained palladium to remove any hydrogen which might be in the gas. The apparatus was evacuated by a quartz mercury pump, and the pressure of the gas was measured by a McLeod gauge. There were two traps cooled by liquid air in the tubing connecting the pump and the gauge with the quartz bulb, so that no mercury vapour could diffuse into the bulb. With these and other similar precautions, there was every reason to believe that the gas used was quite pure.

The oscillating force was maintained between two plates, P_1P_2, (fig. 1), on either side of the bulb B, which formed the condenser of a secondary circuit loosely coupled to a high-frequency continuous wave generator. The latter was mounted on wheels, and by its movement the coupling between the generator and the secondary circuit could be adjusted. The current in the line XY was determined by means of the small aperiodic circuit A connected through the vacuo-junction J to the galvanometer G. The sensitivity of this measuring apparatus could be varied by changing the distance of the circuit A from the line XY.

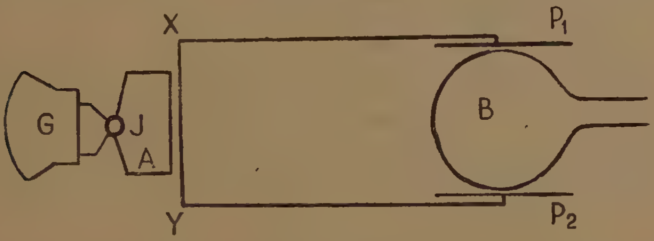
---

* Communicated by Professor J. S. Townsend, F.R.S.

† Phil. Mag. 7, xxvi. p. 290 (Aug. 1938).

The electric force between the plates P_1 and P_2 is proportional to the current in circuit A, and by the use of miniature valves the galvanometer was calibrated to give the potential between the plates in volts. This calibration was not very accurate as the frequency of the oscillations was so high, but the method served to give an approximate estimate of the electric forces on the gas in B.

Fig. 1.



A magnetic force at right-angles to the direction of the high-frequency electric force was maintained in the space between the plates by a large pair of Helmholtz coils.

2. In the theory of the oscillatory discharges which was given, the following results were obtained :—

(*a*) The force required to start a discharge at high pressures of the gas is independent of the frequency.

(*b*) The force required to start a discharge at low pressures increases with the frequency.

(*c*) For each frequency there is a small range of values of the magnetic force, in which the magnetic force causes a *reduction* in the oscillating force necessary to start a discharge in the gas at a low pressure. The reduction is greatest when the relation between the frequency $q/2\pi$ of the force $X \sin(qt+\alpha)$ and the magnetic force H is $H=qm/e$, e being the charge and m the mass of the electron.

(*d*) The force required to maintain a discharge is considerably less than that required to start it.

3. Briefly, the reasons for the above four conclusions are as follows :—

(*a*) For a steady state, the gain of energy of an electron due to its movement in the direction of the electric force may be equated to the loss of energy by collision, and the expression for the mean energy becomes

$$E(1+q^2T^2)^{\frac{1}{2}}=Xel/2\sqrt{\lambda}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

λE being the mean loss of energy of an electron in a collision, l the mean free path of an electron, and T the mean interval between collisions. The time T decreases as the pressure rises, and for those values of the pressure such that qT is less than, say, 0·3, q^2T^2 may be neglected in the expression $(1+q^2T^2)$, and the expression for E becomes

$$E=Xel/2\sqrt{\lambda}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

which is independent of the frequency of the oscillations, and it follows from the theory that the starting potentials are also independent of the frequencies at high pressures.

(*b*) At low pressures T increases, and for values of qT greater than, say, 3, $(1+q^2T^2)$ may be written as q^2T^2, when equation (1) becomes $EqT=Xel/2\sqrt{\lambda}$, and eliminating l/T by the relation $E=ml^2/2T^2$, the equation for the mean energy becomes

$$E=X^2e^2/2mq^2\lambda. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

It follows that at pressures corresponding to values of qT greater than 3, the force required to start a discharge is proportional to the frequency q.

(*c*) When a magnetic force acts on the electrons in a direction perpendicular to the electric force, the mean energy of agitation of the electrons may easily be calculated in the two cases where the phase α of the oscillatory force at the time $t=0$ is zero or $\pi/2$.

In the former case the force is X sin (qt), and if H is the strength of the magnetic field acting along the axis y, then, if u and w are the velocities of the electron along the axes x and z, the accelerations are given by :

$$m\frac{du}{dt}=Xe\sin qt+Hew,$$

$$m\frac{dw}{dt}=-Hew,$$

and writing Xe/m as f, the expressions for u and w become

$$u=w_0\sin\omega t+u_0\cos\omega t+fq(\cos qt-\cos\omega t)/(\omega^2-q^2),$$

$$w=w_0\cos\omega t-u_0\sin\omega t+f(q\sin\omega t-\sin qt)/(\omega^2-q^2),$$

where $\omega=He/m$.

"Resonance" occurs when q is equal to ω, and in that case, by simplification, u and w may be expressed as follows :—

$$u=(w_0+ft/2)\sin\omega t+u_0\cos\omega t,$$
$$w=(w_0+ft/2)\cos\omega t-u_0\sin\omega t-(f/2\omega)\sin\omega t.$$

For values of t greater than π/ω, the energy at any time t is given by

$$m(u^2+w^2)/2=m[(w_0+ft/2)^2+u_0^2]/2,$$

which is the energy acquired by an electron under the action of a direct force of intensity X/2 in the direction of the axis z.

In the case where the phase at time $t=0$ is $\pi/2$, the energy at "resonance" is found to be equal to that which would be acquired under the action of a direct electric force X/2 along the axis x. And, in general, when the mean interval between collisions is greater than half the periodic time of the oscillatory force, the kinetic energy acquired by the electrons under the electric force X sin $(qt+\alpha)$ and the magnetic force qm/e is equal to the energy which would be acquired by the electrons under the action of a direct force of intensity X/2 and no magnetic field.

But, from equation (1), the energy acquired by an electron under the oscillatory force alone is $\mathrm{X}el/2\sqrt{\lambda(1+q^2\mathrm{T}^2)}$, which is the energy which would be acquired under a direct electric force of intensity $\mathrm{X}/\sqrt{2(1+q^2\mathrm{T}^2)}$. The magnetic field adjusted to "resonance" should therefore, at low pressures, reduce the value of the starting potentials.

(*d*) As the energy of agitation of the electrons is increased by increasing the electric force, a point is reached when the rate at which electrons are generated in the gas is equal to the rate at which they are lost by diffusion to the walls of the bulb. At this point a glow appears, and a large number of electrons is produced in the gas. The electrons tend to move to the walls more quickly than the positive ions which are produced with them ; hence, when a discharge occurs, there accumulates in the gas a positive charge, which retards the motion of the electrons to the walls. Under these conditions, the rate of production of electrons equal to the rate of loss by diffusion is maintained by a smaller electric force. That is, the force necessary to maintain a discharge is less than that required to start it.

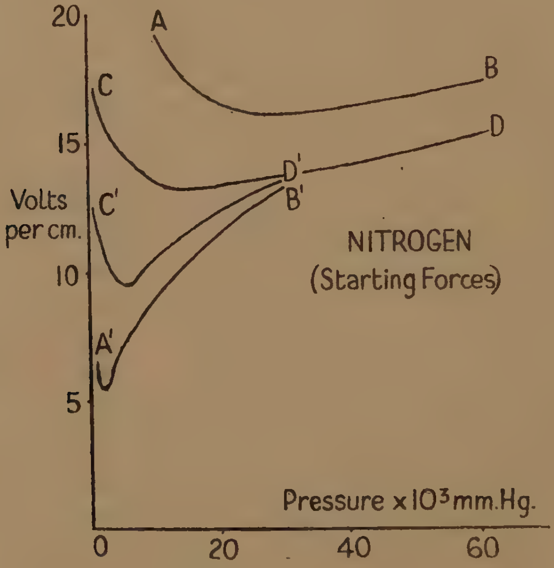
4. The extent to which the various conclusions arrived at have been verified experimentally is shown by the curves in fig. 2.

The wave-lengths of the oscillations were 5·9 and 9·2 metres. The curve AB gives the amplitude of the forces of wave-length 5·9 metres in volts per centimetre necessary to start a discharge, in terms of the pressure of the gas ; the curve CD gives the amplitude of the forces of wave-length 9·2 metres. The reading on the galvanometer (which for a fixed frequency is accurately proportional to the square of the amplitude) was taken at the instant that a faint glow appeared in the gas.

The theory indicates that the curves should meet at a pressure greater than 1 mm. of mercury. With the generator available it was impossible to produce a discharge at this high pressure at either of the frequencies.

The curves show that the force required to start a discharge at low pressures increases with the frequency.

Fig. 2.



5. The curves A′B′ and C′D′, fig. 2, give the minimum value of the electric force required to start a discharge in a magnetic field. For each pressure, measurements were made of the starting potentials with magnetic forces of different intensities. Typical results are given in fig. 3. The curves show that for a certain value of the magnetic force the starting potential is a minimum; and further, the extent to which the efficiency of the magnetic field in reducing the starting potential depends on the pressure may be seen.

It was found, in a sample run, that for the 9·2 metre wave, whereas the expression $H=qm/e$ gives a value of H of 11·6 gauss, the observed value was 11·5 gauss. In the case of the 5·9 metre wave, the two values were 18·1 and 18·2 gauss *.

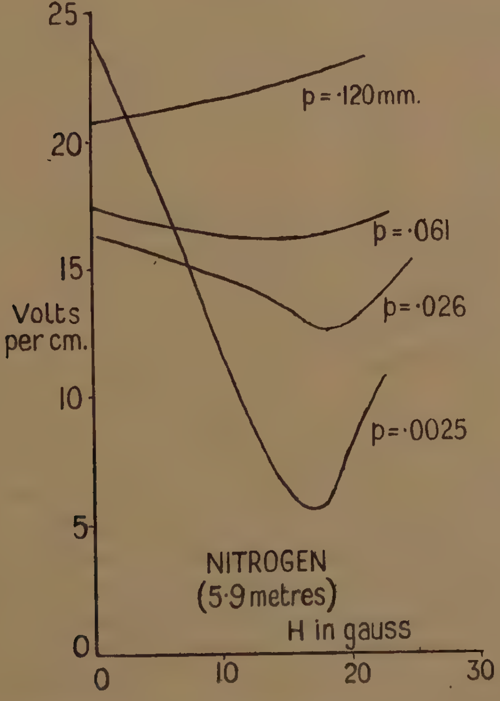
---

* Although experimental difficulties were too great to allow a full investigation of a 3·6 metre wave, sufficient results were obtained to indicate that here, too, "resonance" occurred when He/m was equal to q.

It is interesting to note that, with the generator available, it was impossible, at the lowest pressures, to produce any discharge in the absence of the " resonance " magnetic field.

6. We have seen that the energy of agitation of the electron under the influence of the resonance field is independent of the frequency, being, in fact, the energy corresponding to a direct force $X_h/2$. But H, which is in this case qm/e, is proportional to the frequency. The rate of diffusion of the electrons diminishes as H increases, that is, as the frequency increases. Accordingly, we should expect a reduction in the essential rate of ionization which would manifest itself as a reduction in the value of X_h with increasing frequency at low pressures.

Fig. 3.



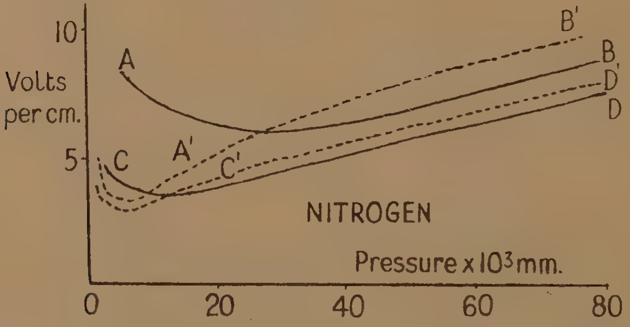
This anticipated result of the influence of the magnetic field on diffusion is found, since it is seen from fig. 2 that at low pressures the " magnetic curve " A′B′ for 5·9 metres lies wholly below the corresponding 9·2 metre curve C′D′. (Whereas AB lies wholly above CD.)

7. The curves AB and CD in fig. 4 represent the minimum potentials necessary to maintain the discharge in the gas with no magnetic force

acting (AB for 5·9 metres and CD for 9·2 metres). The curves A′B′ and C′D′ represent the potentials, when a magnetic force H is acting and the force is adjusted to the value $H = qm/e$.

It is seen by comparing the ordinates of curves AB and CD of fig. 2 with those of curves AB and CD in fig. 4 that the potential required to maintain the discharge is considerably less than the potential required to start it; for example, with the oscillations of wave-length 9·2 metres, at a pressure of 0·012 mm., the two potentials are 13·3 and 3·3 volts. The effect of the magnetic force in reducing the electric potential required to maintain a discharge is much less than its effect on the electric force necessary to start a discharge. The effect to be expected depends principally on the change in the coefficient of ionization α' due to a change in the value of X/p. By comparison of the force required to produce a discharge (fig. 2) with the force required to maintain the discharge at the same pressure (fig. 4), it may be seen that in the latter case X/p has already been reduced to less than one-third of its value in the former. Any further reduction in X/p consequent upon a reduction of X by the application of the magnetic force would therefore reduce α' to a value so small as to be inconsistent with the production of a discharge. At the lowest pressures, however, X/p increases again to a value which allows of a reduction in α', but to a value not so low as to forbid the occurrence of a discharge.

Fig. 4.



8. An exactly similar series of experiments was carried out using pure helium instead of nitrogen. The gas was prepared from thorianite and the commoner impurities were removed by charcoal cooled in liquid air. Traces of hydrogen were removed by passing a discharge in the gas between two concentric cylinders of oxidized copper gauze, a method which a separate series of experiments showed to be completely effective. As an added precaution, the palladium on the discharge tube was retained, whilst the second side-tube, originally empty, contained cooled charcoal.

9. The results obtained using helium and wave-lengths of 5·9 metres and 9·2 metres are shown in figs. 5 and 6 (starting and maintenance forces). The curves show exactly the same features as were found in the case of nitrogen, but with helium the forces involved are smaller. As in the case of nitrogen, the anticipated relation between the frequency and the optimum magnetic field was observed.

Fig. 5.

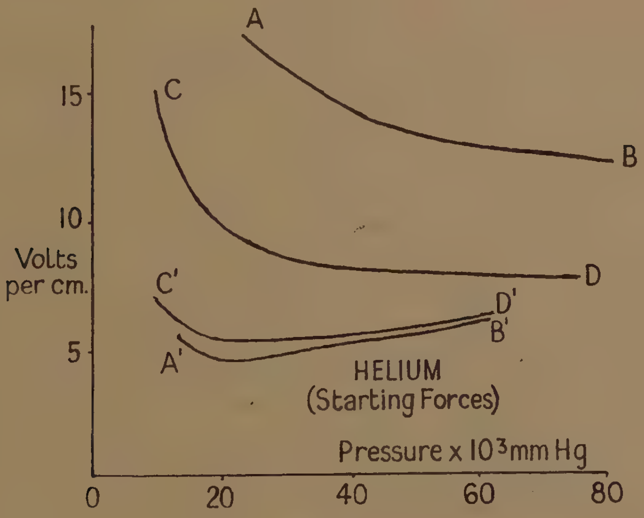
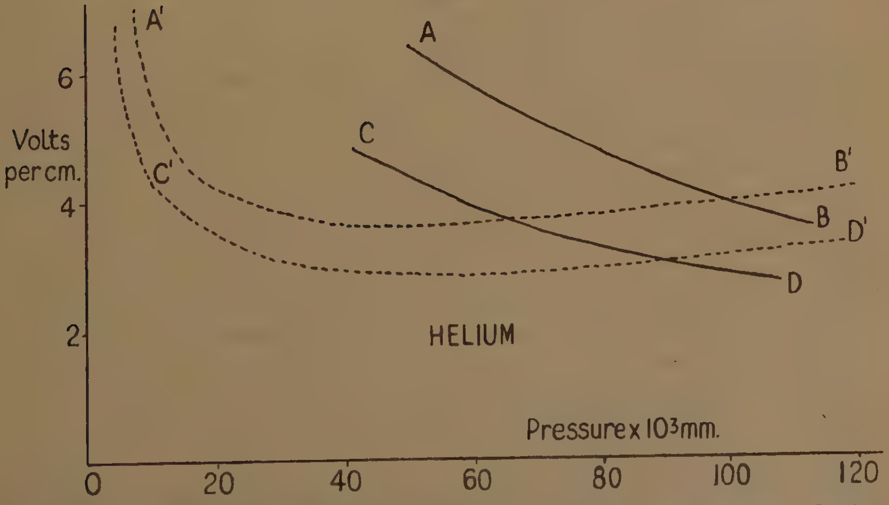


Fig. 6.



I wish to thank Professor Townsend for his constant advice during the progress of this work.

Electrical Laboratory,
Oxford.

XXXII. *Notices respecting New Books.*

Probability and Frequency. By H. C. Plummer. [Pp. xi+277.] (London: Macmillan & Co., 1940. Price 15*s*.)

Professor Plummer has performed a valuable service by writing this relatively elementary, yet modern book. There have been several attempts to show biologists or other workers how to use statistical methods without having to study the theory deeply, and other treatises devoted to elucidating the foundations of the subject. This book falls into neither of these categories, but is simply an introduction to the theory and its applications, with the necessary mathematical proofs included.

It opens with a chapter on probability, of the sort that might be found in many algebra books, and points out the two broad bases on which a theory of probability may be founded. The next chapter is on geometrical probability, and is much the best account known to the reviewer. After these preliminaries, the theory of errors is introduced. The normal law is discussed and shown to be equivalent to the choice of the arithmetic mean in the case of a single quantity. It is also shown to follow from the addition of *small* errors of any types, and to be followed by the combined result of a series of errors each of which follows the law. Nevertheless, the author's view of its status may be seen from the following quotation:—" One condition it is imperative that the law should satisfy if it is to have practical value. It must be simple to apply and it must lead to a unique result. Now the normal law of error satisfies this condition in a pre-eminent degree, and to this fact may be attributed its general acceptance, whatever reserve may be entertained with regard to its ultimate truth. It is not necessary to assert that the law will lead in all cases from the observations to the most probable result. It is enough if it will serve as a foundation for a set of rules which it is generally advantageous to follow." In discussing the methods of assessing the precision of the result, Professor Plummer follows a similar train of ideas, and it is interesting to note that measures of error based on powers are considered as a group, the choice of the square resting rather on grounds of convenience, and of not allowing too much weight to outlying observations, rather than on any more fundamental reason. There is a novel suggestion made in connexion with Chauvenet's criterion for the rejection of an observation. The author is inclined to condemn rejection on *a posteriori* grounds, but if it is done, he argues that the observation in question, being an extreme one, is an indication that the mean ought to be moved in its direction; therefore, this indication ought not to be removed without compensation, and so the extreme observation at the opposite end of the series ought also to be rejected.

Chapters 4 and 5 take us away from the theory of errors into the subject of statistics. In the former chapter, the reader is introduced to the histogram and the standard deviation. Types of frequency distribution are classified, Pearson's differential equation introduced, and the various types of solution considered. The method of moments is well outlined, and Sheppard's corrections are investigated.

The subject of the final chapter is correlation. The treatment is through consideration of lines of regression, and in general follows the Pearson rather than the Scandinavian school. As regards goodness of fit, Pearson's χ test is briefly sketched, but those who propose to use it would probably need to refer to more advanced treatments, of which a selection is given in the form of references at the ends of the chapters. These, by the way, are not necessarily the best, as the author admits. He claims that they will serve to start the student on a course of reading, and that the chain will then extend easily. But if so, is it wise at the end of chapter 4 to give only one reference to a work published this century?

The author's style is particularly pleasant, and the book is well produced. At the end are tables of the gamma and error functions, which add to the usefulness of an already useful treatise. J. H. AWBERY.

A Text Book on Light. By A. W. BARTON, M.A., Ph.D. [Pp. 426+vii.] (Longmans. Price 8*s*.)

GOOD wine needs no bush, and those teachers who already know the author's 'Text Book on Heat' will need no urging to investigate the claims of this book to supply an introduction to the subject suitable for candidates preparing for Higher School Certificate and Scholarship examinations. It is presumably not written for the average schoolboy, for the treatment, though quite clear, is definitely advanced, requiring a sound knowledge of mathematics up to about Higher School Certificate standard and including elementary calculus. Within its self-imposed but very generous limits it is very thorough. Much of the treatment is quite exceptionally excellent; for example, the section on the defects of lens images. The diagrams would be improved by slightly heavier engraving. There are some really beautiful plates of interference and diffraction fringes and the many examples, many of scholarship standard, are well chosen. Altogether a book of good value and one which it is a pleasure to recommend without reserve.

Physics for Medical Students. A supplementary textbook. By J. S. ROGERS, B.A., M.Sc., F.Inst.P. Second Edition. Revised and enlarged. [Pp. xiii +304.] (Melbourne University Press in association with Oxford University Press. Price 12*s*. 6*d*.)

THE purpose of this book is "to supplement the treatment given in standard textbooks." It is certainly true that in spite of the many introductory textbooks, the special interests and requirements of the medical student have been largely ignored to a surprising degree. This book does deal with physical principles and their applications to apparatus and methods of measurement which are of importance to the physiologist and medical man—Osmosis, Colloidal State, Hydrogen Ion Concentration, Blood-Pressure, the Ear, the Eye, the Microscope, Ultraviolet Light, High Frequency Currents, X-rays and Radioactivity, Body Temperature, to mention some of the topics discussed.

The treatment is simple and non-mathematical, but in a number of cases the principles are distinctly skimped, *e. g.* in the sections on the Cardiograph and the Cathode Ray Oscillograph, and there are some surprising omissions.

The account of Colour Vision is somewhat inadequate and there is no mention of the methods actually in use for estimating errors of refraction in the eye. But any medical student who is prepared to pay the price asked will find this a readable and stimulating textbook.

Elements of the Topology of Plane Sets of Points. By M. H. A. Newman. [221 pages.] (Cambridge University Press. Price 12*s.* 6*d.*)

Mr. Newman has written a much-needed introduction to Topology. In doing this he has selected his subject-matter with great care, and the selection is as interesting to the more advanced reader as it is useful to a reader making his first acquaintance with the subject. With an eye no doubt on the great development of Duality Theorems in the last twenty years he has proved Alexander's Lemma in the simplest case—the plane—and then shows to the full its power by the series of theorems developed from it. This occupies roughly half the book : and it is the justification for the book's title. The title is, nevertheless, somewhat misleading, for the remaining half of the book is by no means confined to plane topology. Alexander's Lemma itself and some of its most important consequences are proved also for coordinate spaces of higher dimension, and a large part of the book is devoted to metrical spaces in general. The book is, in fact, more than it claims to be : it is an introduction to the subject as a whole, and in this connexion the "Notes" at the end of the book are excellent for the indications they give for further reading.

The scope of the book may be briefly indicated as follows. Chapter I. deals with that part of the theory of sets of objects which does not involve any geometrical relations between the objects. Chapter II. deals with metric spaces, *i. e.*, sets for which the "distance" between any two members of the set has been defined : it introduces the idea of a limit point and the ideas closed, open, frontier, compact immediately connected with it. Chapter III. is on Homeomorphisms and Continuous Mappings of metric spaces, a homeomorphism being a mapping which is continuous both ways. Chapter IV. is on Connection and includes an excellent section on Local Connection. These four chapters are grouped together as Part I., and in them there is no specialization to the plane or even to coordinate spaces. Chapter V. is entitled "Separation Theorems" and is the most important in the book. After proving Alexander's Lemma, the author develops from it, among other results, Jordan's Theorem, the invariance of dimension number, and the invariance of open sets. The chapter also includes the proof of Alexander's Lemma in higher dimensions and the analogue of Jordan's Theorem in higher space. Chapter VI. is on Simply Connected Domains and includes a theory of Deformations of Paths and a proof of Pollard's form of Cauchy's Theorem. Chapter VII. deals with the Accessibility of points on the frontier of a domain and includes the converse of Jordan's Theorem. The last chapter is entitled "Connectivity Properties," and in it the author gives the theory of the connectivity of open regions, hints at the method of approximation needed for the much deeper theory of the connectivity of closed sets, but rightly confines himself to a very restricted class of closed sets for which the difficulties do not arise. H. D. Ursell.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XXXIII. *The Propagation of Electromagnetic Waves in an Atmosphere containing Free Electrons.*

By L. G. H. Huxley, M.A., D.Phil., F.Inst.P., F. W. Bennett Lecturer in Physics, University College. Leicester *.

[Received April 21, 1939.]

I. *Introduction.*

In a previous paper(1) I have outlined a treatment of the propagation of electromagnetic waves in an ionized atmosphere in which an attempt was made to include a more faithful description of the physical properties of the medium than that suggested by the basic assumptions of the well-known magneto-ionic theory as represented in its most complete development by the fundamental formulæ of Appleton and Hartree(2). In the present paper certain aspects of the subject which were inadequately dealt with in the previous paper will be considered somewhat more fully together with an improved mathematical presentation.

II. *Propagation of Electromagnetic Waves. General Formulæ.*

The following general formulæ which relate the various magnitudes associated with the propagation of a plane electromagnetic wave form the basis of the present treatment, but they may also be conveniently applied to discuss propagation in any medium whatsoever.

We adopt the well-known "cross and dot" notation of vector analysis. In this notation Maxwell's equations for the electromagnetic field are

$$\nabla\times\mathrm{E}=-\frac{\mu}{c}\dot{\mathrm{H}}\,;\ \nabla.\mathrm{E}=4\pi\rho,$$

$$\nabla\times\mathrm{H}=\frac{\dot{\mathrm{E}}+4\pi\mathrm{I}}{c}\,;\ \nabla.\mathrm{H}=0,$$

all electrical quantities being in electrostatic units and H in gauss.

The current density I denotes the vector sum of the current densities of conduction and of polarization and the charge density ρ the sum of the true and "fictitious" $(-\nabla.\mathrm{P})$ charge densities, so that on this understanding it is unnecessary to consider independently the dielectric properties of the medium.

---

*Communicated by the Author.

If the field vectors be assumed to be those associated with a plane electromagnetic wave propagated at phase velocity v, then each may be represented as the product of complex amplitudes containing the phase constants and the common factor

$$e^{jp\left(t-\frac{\mathrm{R}.\vec{n}}{v}\right)}\equiv e^{jp\left(t-\frac{lx+my+nz}{v}\right)},$$

where p is the angular frequency and $\vec{n}$ the unit vector normal to the wave front, the direction cosines of $\vec{n}$ being l, m, and n.

The scalar product $\mathrm{R}.\vec{n}$ is $(lx+my+nz)$, x, y, and z being the coordinates of a point on the wave front and the component magnitudes of the vector R from the origin to that point.

If we define further the polarization vector P through the equation

$$\mathrm{I}=\frac{d\mathrm{P}}{dt}=jp\mathrm{P}.$$

or

$$\frac{d\mathrm{I}}{dt}=-p^2\mathrm{P},$$

the displacement vector D through, $\mathrm{D}=\mathrm{E}+4\pi\mathrm{P}$, and express the scalar product $\vec{n}.\mathrm{E}$ by the symbol Σ, then the following equations may be derived⁽¹⁾ :

$$\left(\frac{c^2}{v^2}-\mu\right)\mathrm{E}=\left(\frac{c^2\Sigma}{v^2}\right)\vec{n}+4\pi\mu\mathrm{P}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\left(\frac{c^2}{v^2}-\mu\right)\mathrm{H}=\frac{4\pi c}{v}(\vec{n}\times\mathrm{P}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\mathrm{H}=\frac{c}{\mu v}(\vec{n}\times\mathrm{E})=\frac{v}{c}(\vec{n}\times\mathrm{D}), \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\frac{c^2}{v^2}\mathrm{E}-\mu\mathrm{D}-\left(\frac{c^2}{v^2}\Sigma\right)\vec{n}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1\,a)$$

$$\vec{n}.\mathrm{D}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

These equations show that in any electromagnetic plane wave, the vectors E, D, and $\vec{n}$ are coplanar [(1 *a*)], the vector H is normal to the plane of E, D, and $\vec{n}$ [(3)], and the vector D normal to that of $\vec{n}$ and H, [(3) and (4)].

Thus the electromagnetic wave is always propagated as transverse waves of D and H.

These relations between the vectors may also be treated as relations between their complex amplitudes as the common factor $e^{jp\left(t-\frac{\mathrm{R}.\vec{n}}{v}\right)}$ may be removed by division.

In what follows the permeability μ will be given the value unity, and the following equations, which are identical with, or are modifications of, the above will be specially important. On writing

$$Q=(c^2q^2-1)=\frac{1}{Q'}, \quad \text{where} \quad q=\frac{1}{v},$$

we have $[\mu=1]$

$$\vec{n}\times(QE-4\pi P)=0=\vec{n}\times\left(\frac{E}{4\pi}-Q'P\right), \quad . \quad . \quad . \quad . \quad (5)$$

$$\vec{n}.D=\vec{n}.(E+4\pi P)=0=\vec{n}.\left(\frac{E}{4\pi}+P\right), \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$H=cq(\vec{n}\times E)=4\pi cqQ'(\vec{n}\times P). \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

III. *Summary of the Usual Magneto-ionic Theory.*

In order the better to compare the usual theory with the alternative treatment given later, a brief summary of the magneto-ionic theory as generalized by Appleton and Hartree from the earlier treatments of Lorentz and others will first be given.

It has been found convenient, however, to refer the phenomena to a system of coordinates in which the Oz axis is along the applied magnetic field H_0 with the plane xOy normal to this direction, and not to one, as in the usual treatments (2) and (3), in which the axis Ox is taken to be the direction of propagation. This latter system is a highly unsuitable one for the discussion of wave propagation by the alternative method, nor is its adoption in accordance with the usual practice in the optics of media which are not isotropic.

The method adopted is to investigate the motion of a single ion under the action of the electromagnetic wave and to obtain as the expression for the current density, the product of the velocity of a single ion, its charge and the number of ions per unit volume, on the assumption that all ions move identically.

Adopting the usual notation[(2), (3)], the equations of motion of an ion in a system of coordinates in which the applied field H_0 is along Oz, and the direction cosines of the wave normal n are l, 0, and n (the symmetry of the problem permitting the simplification of considering $\vec{n}$ to lie in the xOz plane without loss of generality) are

$$\left.\begin{aligned} m\dot{u}&=e(E_x+4\pi hP_x)-fx-gu+\frac{H_0ev}{c}, \\ m\dot{v}&=e(E_y+4\pi hP_y)-fy-gv-\frac{H_0ev}{c}, \\ mw&=e(E_z+4\pi hP_z)-fz-gw, \end{aligned}\right\}, \quad . \quad . \quad . \quad (8)$$

where the ionic charge e is considered positive, and fx etc. are the components of a quasi-elastic binding force, gu etc. are frictional retarding forces, and $4\pi h P_x$ etc. the components of the Lorentz correction term (10) to the field, h usually being taken equal to $\frac{1}{3}$, and m the mass of the ion.

If N ions are present per cubic centimetre of the medium then the components of the current density are

$$I_x = Ne\frac{dx}{dt}, \text{ etc.},$$

whence

$$I_x = jpNex, \text{ etc.},$$

$$P_x = Nex, \text{ etc.},$$

$$x = \frac{P_x}{Ne}; \quad \frac{dx}{dt} = u = \frac{jpP_x}{Ne} \text{ etc.}; \quad \dot{u} = -\frac{p^2P_x}{Ne}, \text{ etc.}$$

Equations (8) assume therefore the following form, in which

$$\omega = \frac{H_0 e}{mc},$$

$$\left.\begin{aligned} \frac{E_x}{4\pi} &= (\alpha + j\beta)P_x - j\gamma P_y = aP_x - bP_y, \\ \frac{E_y}{4\pi} &= (\alpha + j\beta)P_y + j\gamma P_x = aP_y + bP_x, \\ \frac{E_z}{4\pi} &= (\alpha + j\beta)P_z \qquad\qquad = aP_z, \end{aligned}\right\} \quad \ldots \ldots \quad (9)$$

where

$$a = \alpha + j\beta; \quad b = j\gamma = j\frac{mp\omega}{4\pi Ne^2},$$

$$\alpha = -\frac{p^2m}{4\pi Ne^2} - h + \frac{f}{4\pi Ne^2} = -\frac{p^2}{p_0^2} - h + \frac{f}{mp_0^2},$$

$$\beta = \frac{pg}{4\pi Ne^2} = \frac{pg}{mp_0^2},$$

$$\gamma = \frac{p\omega}{p_0^2}; \quad p_0^2 = \frac{4\pi Ne^2}{m}.$$

The components H_{0T} and H_{0L} of the applied field along and normal to the direction of propagation are respectively lH_0 and nH_0. They serve to define the following quantities:

$$p_L = n\omega; \quad p_T = l\omega; \quad \gamma_L = \frac{pp_L}{p_0^2}; \quad \gamma_T = \frac{pp_T}{p_0^2}.$$

It is required to calculate the complex refractive index of the medium and the state of polarization of the wave from equations (9) and the electromagnetic equations of Maxwell. This calculation may be accomplished by the following convenient method.

Equations (5) and (6) taken in conjunction with equations (9) give the following three independent equations :—

$$\left.\begin{aligned} &bP_x-(Q'-a)P_y &&=0, \\ &n(Q'-a)P_x+nb\,P_y-l(Q'-a)P_z&&=0, \\ &l(a+1)P_x-lb\,P_y+n(a+1)P_z&&=0. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Eliminate $\frac{P_x}{P_z}$ and $\frac{P_y}{P_z}$ to obtain

$$\begin{vmatrix} b & -(Q'-a) & 0 \\ n(Q'-a) & nb & -l(Q'-a) \\ l(a+1) & -lb & n(a+1) \end{vmatrix}=0, \quad . \quad . \quad . \quad (11)$$

or

$$(Q'-a)^2-\frac{l^2b^2}{(a+1)}(Q'-a)+n^2b^2=0, \quad . \quad . \quad . \quad . \quad (11\,a)$$

i. e.,

$$\frac{1}{Q'}=Q=(c^2q^2-1)=\frac{2}{2a+\frac{l^2b^2}{(a+1)}\pm\left(\frac{l^4b^4}{(a+1)^2}-4n^2b^2\right)^{1/2}},$$

or

$$c^2q^2=1+\frac{2}{2(\alpha+j\beta)-\frac{\gamma_T^2}{(1+\alpha+j\beta)}\pm\left(\frac{\gamma_T^4}{(1+\alpha+j\beta)^2}+4\gamma_L^2\right)^{1/2}}, \quad . \quad (12)$$

which is the Appleton-Hartree Formula for the complex refractive index.

The state of polarization in the wave is conveniently defined by the ratio of the components of the wave vector H, respectively normal to, and in, the plane formed by H_0 and the wave normal $\vec{n}$.

The former component is H_y and the latter

$$H'=-nH_x+lH_z.$$

Thus,

$$\Pi=\frac{H_y}{H'}=-\frac{D'}{D_y},$$

where

$$D'=-nD_x+lD_z.$$

From equation (7) we find

$$H'=4\pi cqQ'P_y,$$

$$H_y=4\pi cqQ'(-lP_z+nP_x),$$

or

$$\Pi=\frac{nP_x-lP_z}{P_y}.$$

But from the second of equations (10),

$$\frac{nP_x-lP_z}{P_y}=-\frac{nb}{(Q'-a)},$$

whence

$$\Pi=-\frac{nb}{(\mathrm{Q}'-a)}=\frac{j\gamma_{\mathrm{L}}}{\left(\alpha+j\beta-\frac{1}{c^2q^2-1}\right)}, \quad . \quad . \quad . \quad . \quad (13)$$

which is the usual equation for Π.

It is also convenient to express Π in a form independent of q and involving the direction of propagation only.

We have

$$\Pi=\frac{\mathrm{H}_y}{\mathrm{H}'}=-\frac{\mathrm{D}'}{\mathrm{D}_y},$$

which, in conjunction with equations (6), (7), and (9), give the following three independent equations :—

$$\begin{gathered}n\mathrm{P}_x-\Pi\mathrm{P}_y-l\mathrm{P}_z=0,\\ [\Pi b-n(a+1)]\mathrm{P}_x+[nb+(a+1)\Pi]\mathrm{P}_y+l(a+1)\mathrm{P}_z=0,\\ l(a+1)\mathrm{P}_x-lb\mathrm{P}_y+n(a+1)\mathrm{P}_z=0,\end{gathered}$$

whence

$$\begin{vmatrix} n & -\Pi & -l \\ [\Pi b-n(a+1)] & [(a+1)\Pi+nb] & l(a+1) \\ l(a+1) & -lb & n(a+1) \end{vmatrix}=0,$$

or

$$\Pi^2+\frac{l^2b}{n(a+1)}\Pi+1=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

If the roots of this equation are Π_1 and Π_2, then

$$\begin{gathered}\Pi_1+\Pi_2=-\frac{l^2b}{n(a+1)},\\ \Pi_1\Pi_2=1.\end{gathered}$$

Thus, for any arbitrary direction of propagation $(l, 0, n)$ relative to H_0 the independent and permanent wave forms are two elliptically polarized waves executed in opposite senses such that the ratio of the axes in the one is the inverse of the corresponding ratio in the other.

In the particular case of propagation along the applied field $n=1$ and $l=0$, from (14) $\Pi=\pm j$, or the wave forms are those of two circularly polarized oppositely rotating waves.

For propagation normal to the field $l=1$, $n=0$; $\Pi_1=\infty$, $\Pi_2=0$, representing a pair of plane polarized waves with H respectively normal to and parallel to H_0.

In order to deal with the propagation of an electromagnetic wave through an ionized gas by this method the coefficient f in equations (9) is put equal to zero, and a suitable value is assigned to the coefficient

g, so that the free ions in the gas are, in effect, treated as quasi-elastically bound ions with negligibly weak binding forces.

It can hardly be maintained, however, that equations (9) then give a faithful description of the motion of ions in a gas under the influence of an alternating electric field, for the motion which they represent is more nearly that of an ion free to move in a viscous medium than that of an electron or an ion in a gas.

The only motion represented in equation (9) is the steady motion acquired under the action of the applied field of the wave, whereas in an actual gas the effect of the wave is to superpose on the random agitational velocities U of a group of ions a motion W of the centre of gravity of the group such that W is usually only a few per cent. at the most of the velocities U. It is not sufficient to estimate a value for g by actual consideration of the kinetic theory of the ions in the gas, for the equations (9) still give an imperfect description of the essentially gaseous character of the medium, since the ions actually do not execute identical motions under the action of the applied field.

It is only in the limiting case of vanishing concentration of the gas molecules in which g approaches zero, that equations (9) and the formulæ based on them can be considered rigorous.

For instance, when g is zero β is zero, and the condition for reflexion in the ionosphere is $Q'=-1$ and $a=\alpha$, and equation (11) then reduces to

$$(\alpha+1)\begin{vmatrix} b & (\alpha+1) & 0 \\ -n(\alpha+1) & nb & l \\ l(\alpha+1) & -lb & n \end{vmatrix}=0,$$

or

$$(\alpha+1)[b^2+(\alpha+1)^2]=0,$$

i.e.,

$$\left.\begin{aligned} &\alpha+1=0,\\ \text{or}\qquad &\alpha+1=\pm jb=\mp\frac{mp\omega}{4\pi Ne^2}.\end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \quad (15)$$

These equations are in good agreement⁽²⁾ with phenomena connected with the behaviour of the ordinary and extra-ordinary waves in the F layers of the ionosphere, where collisions between ions and molecules are relatively unimportant.

In those regions of the ionosphere where collisions between ions and molecules must be considered, formulæ based on equations (9) would not be expected to describe the phenomena with equal success as when applied to the F layers.

For instance, phenomena related to the interaction of electromagnetic waves which have been extensively investigated by Bailey⁽⁴⁾ cannot be treated merely on the basis of equations such as (9) without essential modification or extension of these equations by the introduction of new physical hypotheses. In view of the limitations inherent in general treatments of the propagation of electromagnetic waves in ionized gases based on the method of Lorentz, it appeared justifiable to attempt an alternative approach⁽¹⁾ to these problems, in which the actual electrical properties of the medium as a whole are considered from the outset.

On the assumption that the electrical properties of the upper atmosphere in regard to the propagation of radio waves are due to the presence of free electrons, preliminary mathematical investigations were undertaken to obtain a general expression for the conductivity of a gas containing free electrons, the results of which are summarized below.

IV. *Conductivity of a Gas containing Free Electrons.*

From the work of Townsend⁽⁵⁾ and his associates it is known that free electrons in a gas may acquire, under the action of a steady field (or of an alternating field of sufficiently high frequency and constant amplitude), a steady state of motion in which the mean energy of agitation of the electrons may exceed the mean energy of agitation of the molecules of the gas. Thus, if $\frac{1}{2}mU^2$ is the mean energy of agitation of an electron and $\frac{1}{2}m\Omega^2$ that of a molecule of the gas, then $\frac{1}{2}mV^2=k.\frac{1}{2}m\,\Omega^2$, where k is a factor which is never less than unity and is a function of the ratio X/p of the electric field X to the pressure of the gas p.

A group of electrons also drifts through the gas under the action of X with a mean velocity of drift W (the mean velocity of the centre of gravity of the group), where W is also a function of X/p, and is usually only a few per cent. of the velocities U, so that the principal motion of an electron is that of random agitation represented by the velocity U.

The formula⁽⁶⁾ for W in terms of the mean free path l, e/m for the electron, and velocity of agitation U under a steady force X, is

$$W=\frac{2}{3}X\frac{e}{m}.\left(\overline{\frac{l}{U}}\right)=\frac{2}{3}X\frac{e}{m}.\overline{T}, \quad . \quad . \quad . \quad . \quad . \qquad (16)$$

where in calculating $T=\left(\overline{\frac{l}{U}}\right)$ it may be assumed for convenience that l is independent of U in regard to the distribution of velocities about the mean. In what follows the bar sign for the average will usually be omitted for simplicity.

Let a permanent magnetic field H_0 be applied along Oz while the electrons drift through the gas under the action of an electric field X along Ox. The velocity W, no longer parallel to X, has components⁽⁷⁾ ᵃⁿᵈ ⁽⁸⁾.

$$W_x=\frac{2}{3}X\frac{e}{m}.\frac{T}{(1+\omega^2T^2)}\left[1+\frac{\omega^2T^2}{1+\omega^2T^2}\right], \quad . \quad . \quad . \quad . \qquad (17)$$

$$W_y = -\frac{2}{3}X\frac{e}{m}\cdot\frac{\omega T^2}{(1+\omega^2T^2)}\left[\frac{1}{2}+\frac{\omega^2T^2}{1+\omega^2T^2}\right], \quad . \quad . \quad . \quad . \quad (18)$$

so that the deflexion θ of the stream by the magnetic field H_0 is given by

$$\tan\theta = \frac{W_y}{W_x},$$

where e is considered positive, and

$$\omega = \frac{H_0 e}{cm}.$$

In general, under the action of an electric force whose components are X, Y, and Z, with the magnetic field H_0 applied along Oz, the components of W are given by

$$\left.\begin{aligned} W_x &= \frac{2}{3}X\frac{e}{m}\cdot\frac{T}{(1+\omega^2T^2)}\left[1+\frac{\omega^2T^2}{1+\omega^2T^2}\right]+\frac{2}{3}Y\frac{e}{m}\cdot\frac{\omega T^2}{(1+\omega^2T^2)}\left[\frac{1}{2}+\frac{\omega^2T^2}{1+\omega^2T^2}\right], \\ W_y &= \frac{2}{3}Y\frac{e}{m}\cdot\frac{T}{(1+\omega^2T^2)}\left[1+\frac{\omega^2T^2}{1+\omega^2T^2}\right]-\frac{2}{3}X\frac{e}{m}\cdot\frac{\omega T^2}{(1+\omega^2T^2)}\left[\frac{1}{2}+\frac{\omega^2T^2}{1+\omega^2T^2}\right], \\ W_z &= \frac{2}{3}Z\frac{e}{m}T. \end{aligned}\right\} \quad . \quad . \quad . \quad (19)$$

We consider next the formulæ for W due to a simple harmonic force $X = X_0 \cos pt$, first without, and secondly with, an applied magnetic field H_0 along Oz.

Without the field H_0 we find(8) that in general W is not in phase with X, being given by

$$\begin{aligned} W = {} & \frac{2}{3}X_0\frac{e}{m}\frac{T}{(1+p^2T^2)}\left[1+\frac{p^2T^2}{1+p^2T^2}\right]\cos pt \\ & +\frac{2}{3}X_0\frac{e}{m}\cdot\frac{pT^2}{(1+p^2T^2)}\left[\frac{1}{2}+\frac{p^2T^2}{1+p^2T^2}\right]\sin pt, \quad . \quad . \quad . \quad (20) \end{aligned}$$

in which the fluctuation of the mean velocity of agitation during a cycle is neglected, a justifiable procedure in most cases encountered in connexion with the propagation of a steady train wave through the ionosphere.

When H_0 is applied(8) and (1), W is both out of phase with X and is no longer parallel to X.

As the expressions become somewhat cumbersome at this stage it is convenient to introduce the following abbreviations :

$$\left.\begin{aligned} F^{\pm} &= \frac{1}{1+(p\pm\omega)^2T^2}\left[1+\frac{(p\pm\omega)^2T^2}{1+(p\pm\omega)^2T^2}\right], \\ f^{\pm} &= \frac{(p\pm\omega)T}{1+(p\pm\omega)^2T^2}\left[\frac{1}{2}+\frac{(p\pm\omega)^2T^2}{1+(p\pm\omega)^2T^2}\right], \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad (21)$$

$$\left.\begin{aligned} &\mathrm{S}=\mathrm{F}^{+}+\mathrm{F}^{-}; \quad s=f^{+}+f^{-}, \\ &\mathrm{D}=\mathrm{F}^{+}-\mathrm{F}^{-}; \quad d=f^{+}-f^{-}, \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \quad (22)$$

expressions (22) being merely the sums and differences of the functions F and f.

We define also the functions G and g as the forms assumed by F and f respectively when $\omega=0$, which appear in the coefficients in (20).

Then, in terms of these functions,

$$\left.\begin{aligned} \mathrm{W}_x&=\frac{\mathrm{X}_0}{3}\cdot\frac{e}{m}\cdot\mathrm{T}\,[\mathrm{S}\cos pt+s\sin pt], \\ \mathrm{W}_y&=\frac{\mathrm{X}_0}{3}\cdot\frac{e}{m}\cdot\mathrm{T}[\mathrm{D}\sin pt-d\cos pt]. \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \quad (23)$$

When the electric field acts along Oz, H_0 has no effect upon W, which is given by

$$\mathrm{W}_z=\frac{2}{3}\mathrm{Z}_0\frac{e}{m}\mathrm{T}\,[\mathrm{G}\cos pt+g\sin pt]. \quad . \; . \; . \; . \; . \; . \quad (24)$$

Formulæ (23) include the earlier formulæ (16), (17), (18), and (20) as special instances.

We are now in a position to express in a general manner the components of the current density I_x etc. in the gas in terms of the components of the electric wave vector. Adopting the notation of complex quantities and neglecting for simplicity the Lorentz correction term involving the polarization, we have

$$\left.\begin{aligned} \mathrm{I}_x&=\mathrm{N}e\mathrm{W}_x=\frac{\mathrm{N}e^2\mathrm{T}}{3m}\{[\mathrm{S}-js]\mathrm{E}_x+[d+j\mathrm{D}]\mathrm{E}_y\}, \\ \mathrm{I}_y&=\mathrm{N}e\mathrm{W}_y=\frac{\mathrm{N}e^2\mathrm{T}}{3m}\{[\mathrm{S}-js]\mathrm{E}_y-[d+j\mathrm{D}]\mathrm{E}_x\}, \\ \mathrm{I}_z&=\mathrm{N}e\mathrm{W}_z=\frac{2\mathrm{N}e^2\mathrm{T}}{3m}[\mathrm{G}-jg]\mathrm{E}_z. \end{aligned}\right\} \quad . \; . \; . \; . \quad (25)$$

The components of the electric polarization being, by definition,

$$\mathrm{P}_x=\frac{\mathrm{I}_x}{jp}, \text{ etc.,}$$

we may derive the following equations giving the relation between the polarization P and the electric wave vector E :

$$\left.\begin{aligned} 4\pi\mathrm{P}_x&=\mathrm{AE}_x+\mathrm{BE}_y, \\ 4\pi\mathrm{P}_y&=\mathrm{AE}_y-\mathrm{BE}_x, \\ 4\pi\mathrm{P}_z&=\mathrm{CE}_z, \end{aligned}\right\} \quad . \; . \; . \; . \; . \; . \; . \quad (26)$$

where
$$\left.\begin{aligned} A &= -\frac{4\pi N e^2 T}{3mp}[s+jS], \\ B &= \frac{4\pi N e^2 T}{3mp}[D-jd], \\ C &= -\frac{8\pi N e^2 T}{3mp}[g+jG]. \end{aligned}\right\} \quad \ldots\ldots\ldots \quad (27)$$

V. *The Propagation of Electromagnetic Waves in an Atmosphere containing Free Electrons.*

When the equations, such as (9) or (26), which relate the vectors P and E, for the medium, are given, then the characteristics of the wave propagation may be immediately obtained from equations (1) to (4).

From the relations (5), (6), and (26) we obtain the following three independent equations :

$$\left.\begin{aligned} BE_x+(Q-A)E_y &= 0, \\ n(Q-A)E_x-nBE_y-l(Q-C)E_z &= 0, \\ l(A+1)E_x+lBE_y+n(C+1)E_z &= 0, \end{aligned}\right\} \quad \ldots\ldots \quad (28)$$

whence, on eliminating E_x, E_y and E_z,

$$\begin{vmatrix} B & (Q-A) & 0 \\ n(Q-A) & -nB & -l(Q-C) \\ l(A+1) & lB & n(C+1) \end{vmatrix} = 0, \quad . \quad (29)$$

or

$$Q^2 - \frac{(n^2+1)(C+1)A+l^2(A^2+B^2+C)}{l^2A+n^2C+1}Q + \frac{(n^2+C)(A^2+B^2)+l^2AC}{l^2A+n^2C+1} = 0, \quad \ldots \quad (29\,a)$$

where
$$Q=(c^2q^2-1).$$

This equation is more complicated than the corresponding equation (11 *a*) owing to the fact that in the final of equations (26) the coefficient is not the same as either of those in the first two, as is the case in equations (9). The assumption of an identical motion for all ions produces a false simplification in the first treatment.

Although the expression for the complex refractive indices obtained from the roots of (29 *a*) are very complicated in the general case, simple formulæ may be obtained in special instances.

For propagation along the applied field H_0, we have $l=0$, $n=1$, and equation (29) becomes

$$\begin{vmatrix} B & (Q-A) & 0 \\ (Q-A) & -B & 0 \\ 0 & 0 & (C+1) \end{vmatrix} = 0,$$

or

$$(C+1)[(Q-A)^2+B^2]=0,$$

i. e.,

$$Q=A\pm jB=(c^2q^2-1).$$

For propagation at right angles to H_0 we have $l=1$, $n=0$, and (29) reduces to

$$\begin{vmatrix} B & (Q-A) & 0 \\ 0 & 0 & -(Q-C) \\ A+1 & B & 0 \end{vmatrix}=0,$$

or

$$(Q-C)[B^2-(A+1)(Q-A)]=0,$$

i. e.,

$$Q=C$$

or

$$Q=A+\frac{B^2}{(A+1)}.$$

The state of polarization is expressed, as before, through the symbol $\varPi$ defined by

$$\varPi=\frac{H_y}{H'}=-\frac{D'}{D_y},$$

where

$$H'=-nH_z+lH_x,$$

$$D'=-nD_x+lD_z.$$

From formulæ (7) and (6) we have

$$\left.\begin{array}{llll} nE_x & -\varPi E_y & -lE_z & =0, \\ {[\varPi B+n(A+1)]E_x} & +[nB-\varPi(A+1)]E_y & -l(C+1)E_z & =0, \\ l(A+1)E_x & +lBE_y & +n(C+1)E_z & =0, \end{array}\right\} \quad . \quad (30)$$

which give

$$\begin{vmatrix} [\varPi B+n(A+1)] & [nB-\varPi(A+1)] & -l(C+1) \\ n & -\varPi & -l \\ l(A+1) & lB & n(C+1) \end{vmatrix}=0, \quad (31)$$

or

$$\varPi^2-\frac{l^2[B^2+(A+1)(A-C)]}{nB(C+1)}\varPi+1=0. \quad . \quad . \quad . \quad . \quad (31\,a)$$

The products of the roots $\varPi_1\varPi_2=1$, so that, in general, we have two oppositely rotating elliptically polarized waves, the ratio in the axes in the one being the inverse of that in the other.

In the special case of propagation along H_0 we have $l=0$, $n=1$, giving $\varPi=\pm j$, and for propagation normal to the field $l=1$, $n=0$.

$$\varPi_1=\infty,$$

$$\varPi_2=0,$$

representing two waves plane polarized, one parallel to H_0 and the other normal to it.

The quantity Π may also be expressed in terms of Q if we replace one of the three equations (30) by one of the equations (28) involving Q, thus :

$$\left.\begin{array}{lll} nE_x & -\Pi E_y-lE_z & =0, \\ BE_x & +(Q-A)E_y & =0, \\ \multicolumn{3}{l}{l(A+1)E_x+lBE_y+n(C+1)E_z=0,} \end{array}\right\} \quad \ldots\ldots\ldots \quad (32)$$

whence

$$\begin{vmatrix} n & -\Pi & -l \\ B & (Q-A) & 0 \\ l(A+1) & lB & n(C+1) \end{vmatrix}=0, \quad .\ . \quad (33)$$

or

$$\Pi=\frac{l^2B-(Q-A)(l^2A+n^2C+1)}{nB(C+1)};$$

alternatively we may use the following three equations :—

$$\left.\begin{array}{llll} nE_x & -\Pi E_y & -lE_z & =0, \\ n(Q-A)E_x & -nBE_y & -l(Q-C)E_z & =0, \\ l(A+1)E_x & +lBE_y & +n(C+1)E_z & =0, \end{array}\right\} \quad .\ . \quad (34)$$

whence

$$\begin{vmatrix} n & -\Pi & -l \\ n(Q-A) & -nB & -l(Q-C) \\ n(A+1) & lB & n(C+1) \end{vmatrix}=0, \quad .\ .\ .\ . \quad (35)$$

or

$$\Pi=\frac{nB(C+1)}{l^2(Q-C)(A+1)+n^2(Q-A)(C+1)}, \quad .\ .\ .\ .\ . \quad (35\,a)$$

a form corresponding more closely to (13).

VI. *Comparison of the Two Treatments.*

In order to compare the two treatments, we transform equations (26) as follows :—

$$\left.\begin{array}{l} \dfrac{E_x}{4\pi}=a'P_x-b'P_y, \\ \dfrac{E_y}{4\pi}=b'P_x+a'P_y, \\ \dfrac{E_z}{4\pi}=c'P_z, \end{array}\right\} \quad \ldots\ldots\ldots \quad (26\,a)$$

where

$$a'=\frac{A}{A^2+B^2}; \quad b'=\frac{B}{A^2+B^2}; \quad c'=\frac{1}{C}.$$

Equations (26 *a*) may now be compared directly with the corresponding equations (9).

It is easy to see that in the special case of propagation in the F region of the ionosphere where collisions between electrons and gas molecules may be neglected, the two treatments lead to identical results.

In this case we have

$$(p\pm\omega)\mathrm{T}\gg 1$$

$$\mathrm{F}^{\pm}\rightarrow 0\;;\quad f^{\pm}\rightarrow\frac{3}{2(p\pm\omega)\mathrm{T}}\,;\quad g\rightarrow\frac{3}{2p\mathrm{T}}\,,$$

$$\mathrm{S}=\mathrm{D}=0.$$

$$s\rightarrow\frac{3p}{(p^2-\omega^2)\mathrm{T}}\,;\quad d\rightarrow-\frac{3\omega}{(p^2-\omega^2)\mathrm{T}}\,,$$

$$\mathrm{A}=-\frac{4\pi\mathrm{N}e^2}{m(p^2-\omega^2)}\,;\quad \mathrm{B}=j.\frac{4\pi\mathrm{N}e^2}{m(p^2-\omega^2)}\cdot\frac{\omega}{p}\,;\quad \mathrm{C}=-\frac{4\pi\mathrm{N}e^2}{mp}\,,$$

whence

$$a'=-\frac{mp^2}{4\pi\mathrm{N}e^2}\,;\quad b'=j.\frac{m\omega p}{4\pi\mathrm{N}e^2}\,;\quad c'=-\frac{mp^2}{4\pi\mathrm{N}e^2}=a'.$$

Reference to equation (26) shows that in these circumstances ($g=0$) we have (apart from the Lorentz factor h), $a'=a=c'$; $b'=b$.

Consequently, the two treatments here yield identical formulæ. For instance, the equations corresponding to (15) relating to the conditions for reflexion of waves from the ionosphere are

$$(\mathrm{C}+1)=0\quad\text{and}\quad\mathrm{B}^2+(\mathrm{A}+1)^2=0,$$

and are to be derived by substituting $\mathrm{Q}=-1$ in (29). With the special values of A, B and C applicable in this case these formulæ are seen to be identical with (15). The formula for Q and Π in these circumstances are similarly proved to coincide in the two treatments.

If it is desired to introduce the Lorentz factor h into (26 *a*), it can be shown that the coefficients in (26 *a*) then become

$$a'=\left(\frac{\mathrm{A}}{\mathrm{A}^2+\mathrm{B}^2}-h\right);\quad b'=\frac{\mathrm{B}}{\mathrm{A}^2+\mathrm{B}^2}\,;\quad c'=\left(\frac{1}{\mathrm{C}}-h\right).$$

Equations (26) become modified to

$$\left.\begin{aligned}4\pi\mathrm{P}_x&=\frac{[\mathrm{A}-h(\mathrm{A}^2+\mathrm{B}^2)]\mathrm{E}_x+\mathrm{BE}_y}{1-2\mathrm{A}h+h^2(\mathrm{A}^2+\mathrm{B}^2)}\,,\\ 4\pi\mathrm{P}_y&=\frac{-\mathrm{BE}_x+[\mathrm{A}-h(\mathrm{A}^2+\mathrm{B}^2)]\mathrm{E}_y}{1-2\mathrm{A}h+h^2(\mathrm{A}^2+\mathrm{B}^2)}\,,\\ 4\pi\mathrm{P}_z&=\frac{\mathrm{CE}_z}{(1-h\mathrm{C})}\,.\end{aligned}\right\}\quad .\;\;.\;\;.\;\;.\quad(26\,b)$$

Thus, h does not enter as a simple additional term.

Except in the special conditions treated above the coefficients a and b of equations (9) will not, in general, be identical with a' and b' of (26 a), nor will $c'=a'$, and the two treatments will yield divergent formulæ when applied to situations where the collision frequency cannot be neglected.

The second method is suitable for dealing with those phenomena such as the interaction (ordinary or gyro-) of radio waves, which arise from variations in the electromagnetic properties of the medium due to changes in the mean energy of agitation of the electrons in an applied alternating electric field whose amplitude is varied relatively slowly.

VII. *Effect of Distribution of the Velocities of Agitation.*

Except in the limiting condition in which the collision frequency is negligible the distribution function for the velocities of agitation will affect the values of the coefficients A, B, and C.

To illustrate the effect of a changing distribution function on the numerical coefficients in the formulæ for the conductivity of the gas we may consider a simple case discussed by Townsend (9).

From equations (17) and (18) we derive a formula for the deflexion of a stream of electrons by a transverse magnetic field.

When $\omega T \ll 1$ the formula for the deflexion becomes

$$\tan\theta = -\frac{1}{2}\,\frac{\omega\overline{T'^2}}{\overline{T}},$$

a formula obtained independently by Townsend (9).

If we neglect the distribution of velocities U we find

$$\tan\theta = -\frac{1}{2}\,\frac{H_0 e}{mc} = -\frac{3}{4}\left(\frac{H_0}{Xc}\right)\cdot\frac{2}{3}\cdot\frac{Xe}{m}\left(\frac{l}{U}\right)$$

$$= -\frac{3}{4}\left(\frac{H_0}{Xc}\right)W,$$

or

$$\left(\frac{H_0 W}{Xc}\right) = -1{\cdot}33\tan\theta.$$

With a velocity distribution, however, we have

$$\tan\theta = -\frac{1}{2}\left(\frac{H_0 e}{mc}\right)l.\left(\overline{\frac{1}{U^2}}\right)\cdot\frac{1}{\left(\overline{\frac{1}{U}}\right)} = -\frac{3}{4}\left(\frac{HW}{Xc}\right)\cdot\frac{\left(\overline{\frac{1}{U^2}}\right)}{\left(\overline{\frac{1}{U}}\right)^2}.$$

Taking a Maxwellian distribution as a typical wide distribution, we obtain for such a distribution

$$\tan\theta = -\frac{3\pi}{8}\left(\frac{HW}{Xc}\right),$$

or

$$\left(\frac{\mathrm{HW}}{\mathrm{X}c}\right)=\cdot 84 \tan\theta.$$

Thus, a change in the distribution function may make an appreciable change in a numerical coefficient, but to investigate the effects of the distribution of the velocities on the factors A, B, and C would be unjustifiable at this stage, and it is simpler to regard the electrons as endowed with a common velocity of agitation U. In any case the distribution function has no appreciable effect except where the collision frequency of an electron with the molecules of the gas has to be considered.

It has not been possible to discuss the propagation of waves in greater detail, as a great deal of preliminary work would first be required for a general evaluation of the coefficients A, B and C.

We note, however, that, on the basis of equations (8), the Lorentz treatment leads to the following expressions for the components of the current density due to an alternating field $\mathrm{X}=\mathrm{X}_0\cos pt$; $\mathrm{Z}=\mathrm{Z}_0\cos(pt-e)$:—

$$\mathrm{I}_x=\frac{\mathrm{N}e^2\mathrm{T}}{3m}\mathrm{X}_0[\mathrm{S}_\mathrm{L}\cos pt+s_\mathrm{L}\sin pt],$$

$$\mathrm{I}_y=\frac{\mathrm{N}e^2\mathrm{T}}{3m}\mathrm{X}_0[\mathrm{D}_\mathrm{L}\sin pt-d_\mathrm{L}\cos pt],$$

$$\mathrm{I}_z=\frac{2\mathrm{N}e^2\mathrm{T}}{3m}\mathrm{Z}_0[\mathrm{G}_\mathrm{L}\cos(pt-\epsilon)+g_\mathrm{L}\sin(pt-\epsilon)],$$

where

$$\mathrm{S}_\mathrm{L}=\mathrm{F}_\mathrm{L}^{+}+\mathrm{F}_\mathrm{L}^{-}\,;\quad s_\mathrm{L}=f_\mathrm{L}^{+}+f_\mathrm{L}^{-},$$

$$\mathrm{D}_\mathrm{L}=\mathrm{F}_\mathrm{L}^{+}-\mathrm{F}_\mathrm{L}^{-}\,;\quad d_\mathrm{L}=f_\mathrm{L}^{+}-f_\mathrm{L}^{-},$$

$$\mathrm{F}_\mathrm{L}^{\pm}=\frac{3}{2[1+(p'\pm\omega)^2\mathrm{T}^2]}\,;\quad f^{\pm}=\frac{3(p'\pm\omega)\mathrm{T}}{2[1+(p'\pm\omega)^2\mathrm{T}^2]}\,;$$

and $$p'=p+\frac{p_0^2h}{p}.$$

Comparison with (21) shows that, apart from the correction term p_0^2h/p,

$$\mathrm{F}^{\pm}=\frac{2}{3}\left[1+\frac{(p\pm\omega)^2\mathrm{T}^2}{1+(p\pm\omega)^2\mathrm{T}^2}\right]\mathrm{F}_\mathrm{L}^{\pm}\,;\quad \mathrm{G}=\frac{2}{3}\left[1+\frac{p^2\mathrm{T}^2}{1+p^2\mathrm{T}^2}\right]\mathrm{G}_\mathrm{L}.$$

$$f^{\pm}=\frac{2}{3}\left[\frac{1}{2}+\frac{(p\pm\omega)^2\mathrm{T}^2}{1+(p\pm\omega)^2\mathrm{T}^2}\right]f_\mathrm{L}^{\pm}\,;\quad g=\frac{2}{3}\left[\frac{1}{2}+\frac{p^2\mathrm{T}^2}{1+p^2\mathrm{T}^2}\right]g_\mathrm{L}.$$

The absorption of the wave is determined by the quantities $\mathrm{F}^{\pm}$ and G and the refractive indices mainly by $f^{\pm}$ and g.

When $(p\pm\omega)\mathrm{T}$ and $p\mathrm{T}\gg 1$ we have

$$\mathrm{F}\to\frac{4}{3}\mathrm{F}_\mathrm{L}\,;\quad \mathrm{G}\to\frac{4}{3}\mathrm{G}_\mathrm{L}\,;\quad f^{\pm}\to f_\mathrm{L}^{\pm}\,;\quad g\to g_\mathrm{L}.$$

Consequently, the two treatments lead to virtually identical conclusions differing merely in the values obtained for the collision frequency which would be overestimated in these circumstances by about thirty per cent. on the Lorentz treatment.

The divergences are more apparent in situations where $(p \pm \omega)T \geqslant 1$ or near the gyro-frequency where $p \rightarrow \omega$. In the latter circumstance $F \rightarrow 2/3F_L$; $f \rightarrow 1/3f_L$, so that the two treatments would here lead to notably divergent estimates of the refractive index.

In the majority of cases encountered in practice we may conclude that use of the Appleton–Hartree formulæ is not likely to result in serious inaccuracies.

References.

(1) Huxley, L. G. H., "The Propagation of Electromagnetic Waves in an Ionized Atmosphere," Phil. Mag. xxx. p. 148 (Jan. 1938).
(2) Appleton, E. V., "Radio Exploration of Upper Atmospheric Ionization," 'Physical Society Reports on Progress in Physics,' ii. p. 129.
(3) White, F. W. G., 'Electromagnetic Waves,' chap. v. (Methuen Monograph).
(4) Bailey, V. A., "On some Effects caused in the Ionosphere by Electric Waves." Part I. Phil. Mag. xxiii. p. 929 (May 1937). Part II. Phil. Mag. xxxi. p. 425 (Oct. 1938).
(5) Townsend, J. S., 'Motion of Electrons in Gases,' Oxford.
(6) Townsend, J. S., Phil. Mag. xxii. p. 145 (1936).
(7) Huxley, L. G. H., Phil. Mag. xxiii. p. 219 (February 1937).
(8) Huxley, L. G. H., Phil. Mag. xxiii. p. 442 (March 1937).
(9) Townsend, J. S., Phil. Mag. xxiii. p. 880 (May 1937).
(10) Ratcliffe, J. A., Proc. Phys. Soc. li. p. 747 (1939).

XXXIV. *The Theory of the Electric Charge and the Quantum Theory.—* Part I.

By H. T. Flint, Reader in Physics,
University of London, King's College *.

[Received January 15, 1940.]

The theory to be developed here is based upon the principle that the fundamental equation of the quantum theory, which is the quantum equation for an electron in a gravitational and electromagnetic field, can be developed by an appeal to simple geometrical ideas. When the gravitational field is negligible, as in the case of problems considered in the quantum theory, the equation is Dirac's equation. The title suggests that two theories are to be developed, but that is not strictly the case. A theory of the electric charge emerges as the basis of the quantum theory is sought. It will appear that the theory is relativistic and, like the theory of general relativity, it rests on an appeal to geometrical form.

A central feature is the incorporation of the conception of parallel displacement with an accompanying change of length of the vector displaced. This, as is well known, was first introduced by Weyl into theoretical physics for the purpose of including the electromagnetic theory in the theory of relativity by giving a metrical significance to the vector potential. There was never any difficulty in presenting the equations of the electromagnetic theory in the requisite relativistic form ; the difficulty was to see their proper significance in the scheme. E. T. Whittaker described the position very aptly by saying that there was no room for mere actors on the relativistic stage. Everything appearing in the performance had to be part of the stage properties. That is a view which guides the development of the present work. It appears from it that Weyl's idea ought to be introduced for the incorporation of the quantum theory into the general scheme. A place for the electromagnetic theory is sought elsewhere, and it is from this direction that a suggestion comes for the theory of the electric charge. The nature of this union of electromagnetism and gravitation lies at the basis of the theory and will be considered first.

In making an appeal to mathematical form for the development of physical theories the purpose is the same as when an appeal is made

---

* Communicated by the Author.

to a model. The object is to obtain some convenient mode of description for physical phenomena, and it is hoped to see the relation of these to one another. In some cases a unitary theory is sought, and if the attempt meets with success the result is a short cut to a widely embracing notation. The most successful example of an appeal to a model is the example of the atom. In this case an appeal was made to dynamical form. In the example of relativity an elegant and useful notation was obtained by means of an appeal to four-dimensional geometry, from which it appears that physical phenomena may be described as if they took place in four-dimensional space-time.

It is just as profitable to ask whether the space and time of the physical world is in fact a four-dimensional continuum as to enquire whether the atom does in fact consist of a nucleus with electrons in motion in orbits about it. What has been obtained from these conceptions is a useful form of description, and what is required of other newer conceptions is a still more useful notation.

Every notation carries with it some limitation, and an appeal to geometry is subject to this defect as well as any other.

But it can be claimed that a great degree of freedom is associated with the appeal to geometry and that geometrical conceptions correspond to the fundamental concepts of location in use in physics. In the present considerations the theory will indicate some of its own limitations and will suggest a limitation to the possibility of ascribing location in space and time to a point charge.

It indicates a limitation to the conception of localization in ordinary space and time in physics (1).

1. *The Photon and the Electron.*

The first example of the association of the conception of wave and particle is the example of the photon. The expression of this association is by means of the null geodesic of the four-dimensional continuum and by the field equations of the electromagnetic theory.

In the consideration of the particle aspect of the photon by means of the null geodesic it is not necessary to introduce its mass arbitrarily. The photon is the null geodesic. No difficulties concerning the structure and energy of the particle arise.

The case of the electron is in marked contrast. In the space of the theory of relativity the equation of motion in an electromagnetic field is not a geodesic and the mass of the particle is introduced as an extraneous constant. This mass is associated with a certain length ($r_0=e^2/m_0c^2$) related to the mass, which alternatively can be regarded as the extraneous constant. This difficulty was present in the classical theory and stood in the way of the realization of an electromagnetic theory of matter, since the stability of the electron could only be realized by the introduction of non-Maxwellian stresses. In the quantum theory as it is usually presented the charged particle is introduced as a point. This implies the introduction of an infinite energy, a difficulty which may be regarded as inherited from the classical theory since in this theory a point charge would be associated with a divergent energy value. In addition other difficulties with regard to the divergence of the energy value appear which are peculiar to the quantum theory. These difficulties of the quantum theory of the electron appear to indicate a breakdown in our conception of the electron as a charged particle.

These considerations encourage the attempt at a representation of the equation of motion of the electron in the form of a null geodesic.

It is not possible to make this representation in a four-dimensional continuum, but early in the history of general relativity, Kaluza proposed the use of a five-dimensional continuum for the purpose of uniting the theory of gravitation and electromagnetism in a unitary theory (2). The object was to achieve the unity which Weyl sought in the conception of change of length with parallel displacement by another means. In Kaluza's theory the equation of the charged particle is a geodesic but it is not a null geodesic. The first step for our present purpose is to modify Kaluza's theory in order to be able to identify the electron equation with that of a null geodesic (3).

In the theory of relativity the coefficients, (g_{mn}), in the expression for the line element

$$ds^2 = g_{mn} dx^m dx^n , \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where summation over the latin suffixes (m, n) from 1 to 4 is implied, are gravitational potentials.

In Kaluza's theory the coefficients, $(\gamma_{\mu\nu})$, replace the (g_{mn}) in a five-dimensional line element, $d\sigma$, where

$$d\sigma^2 = \gamma_{\mu\nu} dx^\mu dx^\nu . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In this case summation over the greek suffixes (μ, ν) from 1 to 5 is implied. These coefficients are made up of the gravitational potentials and of the electromagnetic vector and scalar potentials according to the following table:

$$\gamma_{mn} = g_{mn} + \alpha^2 \phi_m \phi_n , \quad \gamma_{m5} = \gamma_{5m} = \alpha \phi_m , \quad \gamma_{55} = 1 , \quad . \quad . \quad . \quad (3)$$

where ϕ_m represents a typical component of the four-dimensional electromagnetic potential and where α is a constant which will be referred to as Kaluza's constant.

A value will be assigned to α in order to ensure that the path of an electron may be in the form of a null geodesic.

A table of values of the contravariant components is added for convenience

$$\gamma^{mn} = g^{mn} , \quad \gamma^{m5} = \gamma^{5m} = -\alpha \phi^m , \quad \gamma^{55} = \alpha^2 \phi_m \phi^m + 1 . \quad . \quad . \quad . \quad (4)$$

2. *The Equation of Motion of the Electron* (4).

This equation, as described in classical physics, is

$$\frac{d}{dt}\left(m\frac{dx}{dt}\right)=e\left(\mathrm{E}+\frac{v\times\mathrm{H}}{c}\right)_x \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

for the x-component of the motion.

The equation of the geodesic in the five-dimensional continuum is

$$\frac{d}{d\tau}\left(\gamma_{\mu\nu}\frac{dx^\nu}{d\tau}\right)=\frac{1}{2}\frac{\partial\gamma_{\alpha\beta}}{\partial x^\mu}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where the parameter τ denotes the proper time.

In the absence of a gravitational field the line element in four dimensions takes the form

$$ds^2=dx^2+dy^2+dz^2-c^2dt^2=(dx^1)^2+(dx^2)^2+(dx^3)^2+(dx^4)^2, \quad . \quad (7)$$

where $dx^4=\iota c dt$.

If A and ϕ denote the vector and scalar potential of the electromagnetic theory, the contravariant components, (ϕ^m), are

$$(\phi^1, \phi^2, \phi^3, \phi^4)=(\mathrm{A}_x; \mathrm{A}_y, \mathrm{A}_z, \iota\phi). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The covariant components are

$$(\phi_1, \phi_2, \phi_3, \phi_4)=(\mathrm{A}_x, \mathrm{A}_y, \mathrm{A}_z, \iota\phi). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Thus the components of electric intensity are given by

$$\mathrm{E}_m=\frac{\partial\phi_m}{\partial x^4}-\frac{\partial\phi_4}{dx^m}$$

with $\iota E_x=\mathrm{E}_1$, etc., and the components of magnetic intensity are

$$\mathrm{H}_x=(\mathrm{curl}\ \mathrm{A})_x, \text{ etc.},$$

and may be written

$$\mathrm{H}_1=\frac{\partial\phi_3}{\partial x^2}-\frac{\partial\phi_2}{\partial x^3}, \text{ etc.}$$

Thus the equation (5) may be written in the form

$$\frac{d}{d\tau}\left(\frac{dx^2}{d\tau}\right)=-\frac{e}{m_0c}\left(\frac{\partial\phi_1}{\partial x^m}-\frac{\partial\phi_m}{dx'}\right)\frac{dx^m}{d\tau}, \quad . \quad . \quad . \quad . \quad (10)$$

where the rest mass has been introduced by means of the relation

$$\frac{m}{m_0}=\frac{dt}{d\tau}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The relativistic generalization of equation (10) is

$$\frac{d}{d\tau}\left(g_{mn}\frac{dx^n}{d\tau}\right)-\frac{1}{2}\frac{\partial g_{ab}}{\partial x^m}\frac{dx^a}{d\tau}\frac{dx^b}{d\tau}=-\frac{e}{m_0c}\left(\frac{\partial\phi_m}{\partial x^n}-\frac{\partial\phi_n}{\partial x^m}\right)\frac{dx^n}{d\tau}, \quad . \quad (12)$$

If the case, $\mu = m$, be considered in equation (6), it will be found that, after substituting the values from the equations (3),

$$\frac{d}{d\tau}\left(g_{mn}\frac{dx^n}{d\tau}\right) - \frac{1}{2}\frac{\partial g_{ab}}{dx^m}\frac{dx^a}{d\tau}\frac{dx^b}{d\tau} = -\alpha a_0\left(\frac{\partial\phi_m}{\partial x^n} - \frac{\partial\phi_n}{\partial x^m}\right)\frac{dx^n}{d\tau}, \quad . \quad (13)$$

where $a_0 = \gamma_{\mu 5}\frac{dx^\mu}{d\tau}$, it being assumed that the functions in this equation are independent of x^5.

The case, $\mu = 5$, under these conditions becomes $a_0 =$ constant. Thus, in order that equations (12) and (13) may coincide, it is necessary to write

$$\alpha a_0 = e/m_0 c. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

The geodesic would thus represent the track of any particle of rest mass m_0 carrying a charge e.

The electron will be associated with the null geodesic, e, in this case denoting the charge on the electron. Thus, in this case, in addition to the relation (14),

$$d\sigma^2 = 0.$$

The square of the line element may be written in the form

$$d\sigma^2 = \gamma_{\mu\nu}dx^\mu dx^\nu = g_{mn}dx^m dx^n + (\gamma_{\mu 5}dx^\mu)^2. \quad . \quad . \quad . \quad . \quad (15)$$

Writing

$$d\theta = \gamma_{\mu 5}dx^\mu, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

this takes the form

$$d\sigma^2 = ds^2 + d\theta^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

It will be noted that

$$a_0 = \gamma_{\mu 5}\frac{dx^\mu}{d\tau} = \frac{d\theta}{d\tau}.$$

Thus $d\sigma^2 = 0$ requires

$$d\theta^2 = -ds^2 = c^2 d\tau^2,$$

and consequently

$$a_0 = c.$$

Adopting this value for a_0, the value of Kaluza's constant becomes

$$\alpha = e/m_0 c^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

The constant m_0 is to be identified as the rest mass of the electron, and the expression of the constant α in the form (18) is justified simply by the fact that by means of it the track of the electron takes the form

mass × acceleration = force.

It is to be noted that m_0 enters the theory naturally from the constant α introduced in the coefficients $(\gamma_{\mu\nu})$. It does not enter as the mass of a test particle placed in the field. The position is analogous to that occurring in the case of the photon and the null geodesic is the electron.

The element of length $d\theta$ can be described as the component of $d\sigma$ lying perpendicular to ds, i. e. normal to the space of general relativity.

The electron track makes an angle arc tan $\sqrt{-1}$ with this normal. From equation (16) it follows that $d\theta$ can be expressed in the form

$$d\theta=\alpha\phi_m dx^m+dx^5=c\,d\tau. \quad \quad (19)$$

A further similarity between this theory and that of the photon is evident from the fact that the rest mass of the photon is zero. This results from the fact that the photon is a null geodesic.

The relation between ordinary mass and rest mass is

$$\frac{\mathrm{M}}{\mathrm{M}_0}=\frac{dt}{d\tau}=\frac{\iota c\,dt}{ds},$$

and M_0 vanishes when ds does so.

The generalization of this requires the introduction of a mass, I_0, such that

$$\frac{\mathrm{M}}{\mathrm{I}_0}=\frac{\iota c\,dt}{d\sigma},$$

I_0 vanishing with $d\sigma$.

Applying this to the case of the electron

$$\frac{m}{\mathrm{I}_0}=\frac{\iota c\,dt}{d\sigma},$$

and since

$$\frac{m}{m_0}=\frac{\iota c\,dt}{ds},$$

$$\frac{\mathrm{I}_0}{m_0}=\frac{d\sigma}{ds}. \quad \quad (20)$$

Thus I_0 vanishes with $d\sigma$, though m_0 does not.

I_0, in this case, stands in exact analogy to the rest mass in four dimensions, and its vanishing is required by the principle that the electron is to be treated in five dimensions as the photon is in four.

3. *The Quantization of Electric Charge.*

In order to made the geodesic (13) equivalent to the equation of motion of a charged particle it was seen to be necessary to write

$$\alpha a_0=e/m_0c,$$

where e denotes the charge and m_0 the rest mass of the particle, which need not now be considered to be an electron.

But

$$m_0a_0=m_0\gamma_{5\mu}\frac{dx^\mu}{d\tau},$$

and this means that the fifth component of momentum

$$\pi_5=e/\alpha c. \quad \quad (21)$$

This shows that it is the charge which determines this component and a generalization of the principle of conservation of momentum which makes it apply to the five-dimensional case would mean that the ordinary principle of conservation of momentum, the principle of conservation of energy and the principle of conservation of electric charge are united in one principle.

If the Wilson-Sommerfeld quantum conditions be applied to this fifth component a new condition of quantization results.

To the relation

$$\int p\,dq = nh, \qquad (22)$$

the relation

$$\int \pi_5\, dx^5 = nh \qquad (23)$$

is added.

In order to bring this relation into a form applicable in modern quantum theory it is possible to proceed in a way similar to that which makes (22) become

$$pq - qp = h/2\pi i. \qquad (24)$$

It is first necessary to examine the nature of the fifth coordinate. One of the difficulties of the present theory has been to interpret the fifth component of any vector which appears in the five-dimensional continuum.

In the case of the electron, relation (19) shows how dx^5 is connected with ordinary quantities. It appears that the coordinate x^5 is not so easily interpreted as the fourth component x^4 in the theory of relativity. It may be described as the coordinate conjugate to π_5, and if it be understood that the dimensions are accounted for by the factor $1/\alpha c$, it can be said to be conjugate to the charge.

It has already been assumed that the electromagnetic potential components are independent of x^5, and it appears that this component is absent from many functions occurring in physics. This fact, taken with the result already obtained that the fifth component of momentum is constant, suggests that the fifth coordinate is cyclic in character. It is analogous to the angular coordinate used in the dynamics of atomic orbits. This is often denoted by θ, which is not to be confused with the element $d\theta$ already introduced. The fact that this coordinate does not appear in the energy expression leads at once to the result that the component of momentum, p_θ, is constant and the condition

$$\int p_\theta\, d\theta = nh$$

becomes

$$2\pi p_0 = nh. \qquad (25)$$

In the old quantum theory the actual value of θ had no significance. In the theory of the orbits use was made, as in (25), of the fact that it had a range of 2π and could be said to occur in multiples of this amount. This leads to the suggestion that x^5 is a quantity similar to θ occurring in multiples of a length, l_0, and that values of x^5 within this range are of no significance.

Equation (23) combined with (21) leads to the result

$$\frac{e}{\alpha c}l_0=nh,$$

or

$$e=n.\frac{h\alpha c}{l_0}. \qquad (26)$$

This means that the charge e occurs in multiples of a unit of magnitude $h\alpha c/l_0$(5). Thus the theory leads to the result that electric charges are quantized, occurring in definite quanta of magnitude $h\alpha c/l_0$. Thus the fundamental charge is expressed in terms of constants already introduced in the course of development of the theory. Experiment leaves no doubt that electric charges occur in multiples of the charge of the electron, the magnitude being denoted by e. This symbol is now used in this sense, and we write

$$e=h\alpha c/l_0, \qquad (27)$$

or

$$\alpha=l_0e/hc. \qquad (28)$$

This result may be compared with the value obtained for α, which was introduced on the principle that the electron track is a null geodesic. On comparison with equation (18) it will be found that the two values coincide when l_0 is given the value h/m_0c.

In this way the fundamental length h/m_0c appears in the theory. It is known as the Compton wave-length on account of its occurrence in connexion with the Compton effect.

The functions occurring in the equation of motion of the electron were assumed to be independent of x^5, but when this variable occurs in the functions used in the theory it is found to be sufficient to introduce it in a very simple way. It is only necessary to assume dependence on this variable in the factor $\exp.\left(\frac{2\pi i}{h}\pi_5x^5\right)$ which, in the case of an electron, becomes $\exp.\left(2\pi i\frac{m_0c}{h}x^5\right)$, since $\pi_5=m_0c$ in this case.

4. *The Energy of the Electron.*

The well-known relation between energy and velocity of an electron follows immediately from $d\sigma^2=0$.

For this purpose the values of the coefficients, (g_{mn}), will be assumed to be unity, *i. e.* appropriate to special relativity. The velocity component

$$u_m=\gamma_{m\nu}\frac{dx^\nu}{d\tau}=\gamma_{mn}\frac{dx^n}{d\tau}+\gamma_{m5}\frac{dx^5}{d\tau}$$

$$=g_{mn}\frac{dx^n}{d\tau}+\alpha\phi_m\left(\alpha\phi_n\frac{dx^n}{d\tau}+\frac{dx^5}{d\tau}\right)$$

$$=g_{mn}v^n+\frac{e}{m_0c}\phi_m,$$

$$u_4=g_{4n}v^n+\frac{e}{m_0c}\phi_4=\iota\left(c\frac{dt}{d\tau}+\frac{e}{m_0c}\phi\right).$$

It will also be assumed that the vector potential vanishes, the scalar potential ϕ alone remaining.

Inserting $\frac{dt}{d\tau}=\frac{m}{m_0}$, it follows that

$$m_0cu_4=\iota(mc^2+e\phi). \quad \ldots \ldots \quad (29)$$

Thus the energy of the electron is represented by $-\iota m_0 eu_4$.

5. *Limitations to the Conception of Localization in Space and Time.*

First consider the case when there is no electromagnetic field, *i. e.* each component, (ϕ_m), vanishes.

In this case equation (19) gives

$$cd\tau=dx^5,$$

and, since $dx^5<h/m_0c$ has no physical significance, it follows that

$$d\tau<h/m_0c^2 \text{ }^{(6)} \quad \ldots \ldots \quad (30)$$

has no physical significance.

This is the expression of a principle of minimum proper time. The interpretation is that in the study of the path of an electron it has no meaning to assign to it intervals of proper time less than h/m_0c^2. This applies in the limit to vanishing fields. In order to see how it applies in the case when a field exists, take secondly the case when the vector potential vanishes but a scalar potential exists of the Coulomb type arising from a charge e, i. e. $\phi=e/r$. Equation (19) becomes

$$c\,d\tau=\alpha\phi_4\,dx^4+dx^5$$

$$=-\frac{e}{m_0c}\cdot\frac{e}{r}dt+dx^5.$$

Writing $r_0=e^2/m_0c^2, \quad d\tau=m_0dt/m,$

$$c\left(\frac{m_0}{m}+\frac{r_0}{r}\right)dt=dx^5. \quad \ldots \ldots \quad (31)$$

If the case of a charge e be considered which moves in the field arising from a charge e of opposite sign, this relation becomes

$$c\left(\frac{m_0}{m}-\frac{r_0}{r}\right)dt=dx^5. \quad \ldots \ldots \quad (32)$$

From this the inequality results

$$dt \nless \frac{h/m_0c^2}{m_0/m - r_0/r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

Thus, as r approaches the value mr_0/m_0, the smallest interval of time which can be associated with the path of the charge becomes very large and it is impossible to speak of the particle as localized in time. This is to be interpreted as a breakdown of our representation of electrical phenomena by means of particles localized in space and time. At great distances r, equation (33) gives the result obtained in the absence of a field.

This result may be regarded as showing one aspect of the nature of the difficulty which has led to the impossibility of regarding the nucleus as consisting of unlike charges. The nucleus is a region of linear dimensions e^2/m_0c^2, and is thus an example to which this consideration applies. A similar difficulty does not, according to equation (31), apply to like charges.

It is to be noted that the fundamental length e^2/m_0c^2 enters into this theory naturally without any extraneous idea such as is associated with the concept of electron radius. In this theory the length is rather to be compared with the radius of action such as occurs in the theory of gases. A satisfactory feature of the theory is the inclusion of the two fundamental lengths h/m_0c and e^2/m_0c^2 in a natural way.

6. *Application to Atomic Orbits* (7).

The limitation of the last paragraph has been applied to the case of atomic orbits and it has been shown that, if these orbits be considered as occupied by electrons under a Coulomb field ($\phi = Ze/r$), it is impossible to carry on the well-known description of the Rutherford-Bohr model beyond the number Z=96. If the limitation represents literally a limitation in the physical world of space-time, this result means that the number of chemical elements is limited to 96. If, however, it is merely a limitation included in our method of physical description, it means that we shall have difficulties if we wish to describe any set of observations as arising from atoms built upon this model with a sequence of K, L, etc., layers of electrons of an order higher than 96.

There would apparently be no difficulty if the higher elements lacked K-layers and later on both K and L layers.

In this case the condition suggests the setting in of instability amongst the innermost electrons, and it seems surprising that the condition which determines the atomic number applies only to the electrons. The condition does not impose a limit to the order of the nucleus itself, it is a limitation affecting the order of the atom which consists of a complete set of electron layers. Recent experiments, however, on the rupture of the uranium nucleus indicate that an instability in this structure occurs in that, instead of continuing to build itself up into more massive nuclei, it tends to divide into smaller ones. It seems possible to state that somewhere not far beyond the 92nd element an instability occurs both in the nucleus and in the electrons about it.

Evidence at present appears to point to 93 as the heaviest known atomic number.

7. *The Introduction of other Masses into the Theory.*

The mass denoted by m in the foregoing paragraphs has been described as the mass of the electron. It occurs simply because the null geodesic is thrown into the form

$$\text{mass} \times \text{acceleration} = \text{force}.$$

In this way it appears that the charge e is automatically associated with this mass, which may consequently be described as the electromagnetic mass of the particle, and it is not necessary to associate any mass of another character with it.

But certain phenomena occur in physics which can be described by the device of associating other masses with the charge e. Thus the proton is regarded as a positive charge e associated with a mass M_0.

The motion of the particle is described as before by means of the second law of motion replacing m_0 by M_0.

If the motion is to be described by means of a geodesic, although no longer by a null geodesic, it is necessary to make the identification

$$\alpha \frac{d\theta}{d\tau} = e/M_0 c. \qquad \ldots \ldots \ldots \ldots (34)$$

Thus

$$M_0 \frac{d\theta}{d\tau} = e/\alpha c = m_0 c. \qquad \ldots \ldots \ldots \ldots (35)$$

But since $M_0 \frac{d\theta}{d\tau}$ is the 5th component of momentum, it follows that π_5 has the same value for the proton as for the electron. This is to be expected since this component is determined by the charge carried. The difference in sign of the charge can be referred to a difference in direction of projection of the element, $d\sigma$, along the normal to the four-dimensional space.

It follows, according to the same considerations as before, that for this particle also the minimum proper time is h/m_0c^2.

From the fact that $M_0 \frac{d\theta}{d\tau}$ for the proton is the same as $m_0 \frac{d\theta}{d\tau}$ for the electron, it follows that the projection of the world line element is different in the two cases, the inclination of the proton being arc tan $\left(\iota\frac{M_0}{m_0}\right)$(8).

There are difficulties in understanding the way in which M_0 enters the theory, although in the present description the ratio M_0/m_0 can be interpreted simply as the inclination of the world line to the normal to the four-dimensional space of ds. But this throws no light on the reason why this ratio has a particular value.

The present theory indicates that both for the proton and the electron as also for the positron the operator $\frac{h}{2\pi i}\cdot\frac{\partial}{\partial x^5}$ is equivalent to multiplication by m_0c. Thus x^5 in all these cases may be regarded as occurring in the factor exp. $2\pi i\cdot\frac{m_0c}{h}$.

The introduction of the heavy electron is made by means of a procedure quite different from that of the introduction of the other masses. The mass is of the order of 100 times that of the electron, and this mass is estimated by a consideration of the range of action of the forces between two charges. A certain phenomenon is treated as if it were a case of interaction between two particles separated by a distance of the order of magnitude of the radius of the electron. In this interaction one particle is regarded as a heavy electron moving at great speed in the field of force of the other.

Let it be supposed that an ordinary electron describes a circular path of radius r about a centre of attraction under a field of force with potential $\phi=-e/r$, the centre and electron having of course opposite signs. If the charge of the electron be denoted by e the value of ϕ due to the other charge is $-e/r$.

The interval of time that can be associated with this track is dt, which must satisfy the condition (33),

$$dt=\frac{2\pi r}{v}\not<\frac{h/m_0c^2}{\frac{m_0}{m}-\frac{r_0}{r}}.$$

If β denote the ratio v/c, it follows that

$$r\not<\frac{1}{\sqrt{1-\beta^2}}\left(\frac{h}{2\pi m_0c}\beta+r_0\right). \quad \quad (36)$$

For slow velocities, $\beta\rightarrow0$, the result is the same as that obtained previously, viz. $r\not<r_0$.

It will be noted that the right-hand side consists of two parts, the first, $h\beta/2\pi m_0c$, arises from the fact that x^5 is periodic in character with amplitude h/m_0c. The second arises from the constant α. It appears that this result is true for all the masses described above, where $\pi_5=m_0c=e/\alpha c$ and does not depend on the mass. But in the case of the heavy electron, when β approaches the value unity, r has to be brought down to the order of magnitude of r_0. The only way of doing this is to increase the mass in the first term, and this can be accomplished by assuming for the heavy electron the result

$$\frac{h}{2\pi i}\frac{\partial}{\partial x^5}=\mu_0c, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

instead of
$$\frac{h}{2\pi i}\frac{\partial}{\partial x^5}=m_0c,$$

as for the other particles. Thus π_5 in this case is no longer m_0c but μ_0c, and it is evident that the particle moves according to a different law from that which holds in the case of the other particles.

As a consequence of (37) it is concluded that in functions which concern the theory of the heavy electron x^5 occurs in the factor exp. $(2\pi i\mu_0c/h)$.

In this case it follows that the least distance between this particle and the centre of force is

$$r=\frac{1}{\sqrt{1-\beta^2}}\left(\frac{h}{2\pi\mu_0c}\beta+r_0\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

and, if this is to be of the order of value of r_0, it is necessary that

$$\frac{h\beta}{2\pi\mu_0c}\sim r_0.$$

After substitution of the value of r_0 this leads to

$$\frac{\mu_0}{m_0}\sim\frac{hc\beta}{2\pi e^2}\sim 137\beta.$$

The factor $\frac{1}{\sqrt{1-\beta^2}}$ must not be forgotten in the determination of the value of r, but the value approaches infinity slowly as $\beta\rightarrow 1$. Thus if $\beta=0{\cdot}96$, its value is $3{\cdot}57$.

The conclusion drawn from this is that if certain phenomena are to be explained by the interaction of particles in the way described, it is necessary to introduce one of mass 137β times that of the ordinary electron. For the value of β given above this is a little more than $130m_0$.

In the second part of this theory of the electron field equations will be obtained by the use of the same geometrical background and the same value of α, matrices will be introduced and the quantum equation will be developed.

References.

(1) Flint, Proc. Phys. Soc. l. p. 899 (1938) ; 'Nature,' p. 535, Sept. 17th, 1938.
(2) Kaluza, *S. B. Akad. Wiss. Berlin*, p. 966 (1921).
(3) Fisher, Proc. Roy. Soc. A, cxxiii. p. 489 (1929).
(4) Flint, Proc. Roy. Soc. A, cl. p. 421 (1935).
(5) Flint and Wilson, Proc. Phys. Soc. l. p. 340 (1938).
(6) Flint, Proc. Roy. Soc. A, cxvii. p. 630 (1927).
(7) Flint and Richardson, Proc. Roy. Soc. A, cxvii. p. 637 (1927) ; 'Nature,' p. 535, Sept. 17th, 1938.
(8) O. Klein, *Z. f. Phys.* xxxvi. p. 895 (1926).

XXXV. *Dielectric Constant of Ionized Air.*—III.

By Shah Md. Fazlur Rahman, M.Sc., and S. R. Khastgir, D.Sc.(Edin.), Mem.I.R.E., Dacca University *.

[Received March 24, 1939.]

1. *Introduction.*

In a previous paper it was shown by Gangopadhyaya and Khastgir † that the observed anomalous variation of the effective dielectric constant of ionized air or nitrogen inside a discharge-tube for different values of the tube current was partly, at any rate, due to the formation of a positive ionic sheath round the inner surface of the discharge-tube. In the present paper an account is given of some further experiments on the anomalous dielectric constant of ionized air in a discharge-tube under controlled conditions of tube current and pressure. In measuring the effective dielectric constant of the ionized air, small condenser plates were fixed tightly against the outer surface of the discharge-tube, and measurements were made with an alternating field of a definite ultra-high frequency for a range of pressures from about 2 mm. to about 0·02 mm. of mercury inside the air discharge-tube carrying different values of the current from 0 to 17 m.a. For each different tube current and pressure the electron density and the electron temperature in the region of the experimental condenser were determined by the well-known probe method of Langmuir. The experimental results on the anomalous dielectric constant of ionized air were found consistent with the theory of an ionic sheath formed on the inner surface of the discharge-tube.

2. *Experimental Arrangements.*

The arrangement for the measurement of the effective dielectric constant of the ionized air was the same as in the previous experiments (see fig. 1). A discharge-tube A was fixed in a vertical position between two thin cylindrical metal pieces CC held tightly round the outer surface of the discharge-tube. These pieces constituted the terminal condenser of a pair of Lecher wires. Ultra-high frequency oscillations having been induced in the Lecher system from an oscillator, adjustments for maximum voltage across the terminal condenser were made by sliding a short-circuiting bridge B across the wires. The voltage resonance was indicated by a valve-detecting unit which was coupled to the Lecher system at the input end by two small cylindrical condensers consisting of short brass tubes insulated from the Lecher wires by glass sleeves. The dielectric constant of the ionized air was obtained from

---

* Communicated by the Authors.

† Gangopadhyaya and Khastgir, Phil. Mag. xxv. p. 883 (Suppl., June 1938).

$$\epsilon = \frac{\tan \frac{2\pi l}{\lambda}}{\tan \frac{2\pi l'}{\lambda}}, \quad \ldots \ldots \ldots \quad (1)$$

Fig. 1.

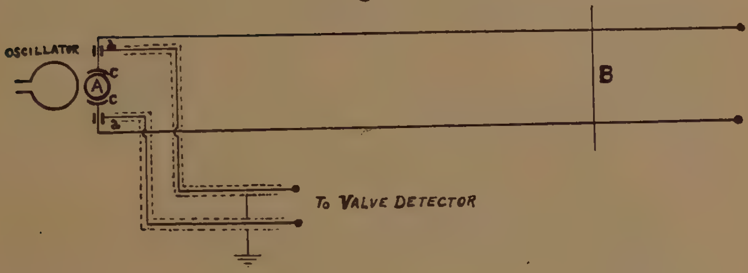
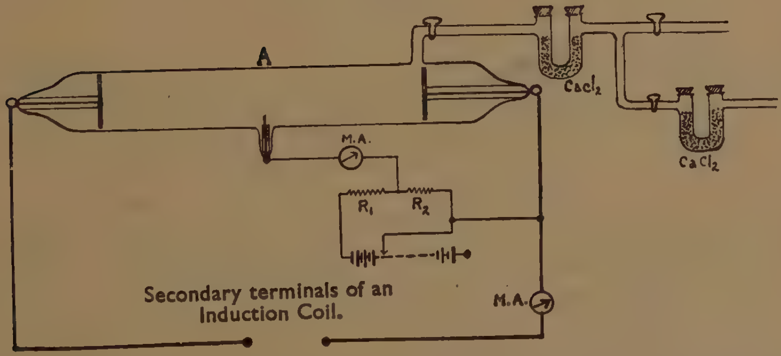


Fig. 2.



where l and l' were the lengths of the Lecher wire measured from the input end up to a point of resonance with the discharge off and on respectively and λ the wave-length of the alternating field.

The electrical connexions for the measurements of the electron density and the electron velocity according to Langmuir's probe method * are shown in fig. 2. A short tungsten wire was inserted in the middle of the discharge-tube round which the cylindrical copper-foil sheets of the experimental condenser were fixed. The wire was sheathed in glass

---

* Langmuir, Gen. Elec. Rev. xxvii. p. 449 (1924) ; Jour. Frank. Inst. cxcvi.

except for a small protruding length which acted as the probe-electrode. A varying negative potential V was applied to the probe-electrode with respect to the anode of the discharge-tube by means of a H.T. battery and a potentiometer arrangement with two high resistances R_1 and R_2 (each of the order of 2 megohms). These resistances joined in series were connected to the H.T. battery of 400–500 volts. One end of one resistance was connected to the anode and the probe was connected to the join of the two resistances. A fraction of the H.T. voltage which was accurately determined was thus applied between the anode and the probe. The probe voltage with respect to the anode was varied by employing different voltage tappings on the H.T. battery. The main discharge was not disturbed by the potentiometer arrangement. The probe current was measured by a suitable milliameter reading fractions of a milliampere.

A Cenco hyvac pump served to exhaust the discharge-tube. The pressure was measured by a McLeod gauge.

The theory of the probe method of measuring the electron velocity and the electron density is briefly given below.

When the applied probe voltage V is less than the space potential V_s, it is well known that the retarded electron current i_- varies with V according to the formula

$$i_-=A.N.e\sqrt{\frac{kT}{2\pi m}}\,e^{\frac{(V-V_s)e}{kT}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where A is the probe area, N the electron density in the discharge, T the electron temperature, k Boltzmann's constant, e and m are the charge and mass of an electron. From this equation we can write

$$\log_e(i_-)=\text{const.}+\frac{(V-V_s)}{kT}e, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

so that when $\log_e(i_-)$ is plotted against V, we should get a straight line of slope given by

$$S=\frac{d(\log_e i_-)}{dV}=\frac{e}{kT}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The temperature of the electrons or kT can be obtained from (4). The average velocity $\bar{v}$ of the electrons can then be obtained from

$$\bar{v}=\sqrt{\frac{8kT}{\pi m}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

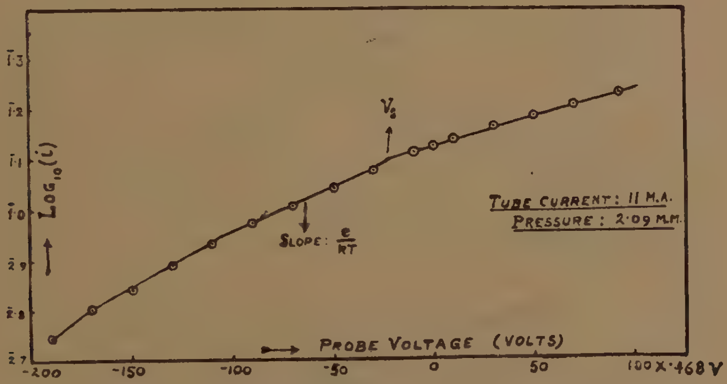
The formula (2) holds good only for retarded electrons, *i. e.*, when $V<V_s$. Hence there is a break in the "semi-log" curve at $V=V_s$. The electron current at this potential of the probe is given by

$$i_-=A.N.e\sqrt{\frac{kT}{2\pi m}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Thus knowing the value of (i_-) at the space potential, N can be calculated from (6) with the help of (4).

In our experiments it was fortunately found that *except for very strongly negative voltages of the probe*, the electron current was sufficiently in excess of the positive ion contribution, so that when the logarithm of the observed total current $(i_- + i_+)$ was plotted against the potential of the probe, there was practically no effect of the positive ion contribution on the straightness of the line, and a discontinuity appeared at $V = V_s$. A typical semi-log curve is shown in fig. 3. The slope of the straight line for $V < V_s$ gave the value of kT. With this value of kT and from a knowledge of the probe current at the discontinuity, N was obtained from (6).

Fig. 3.

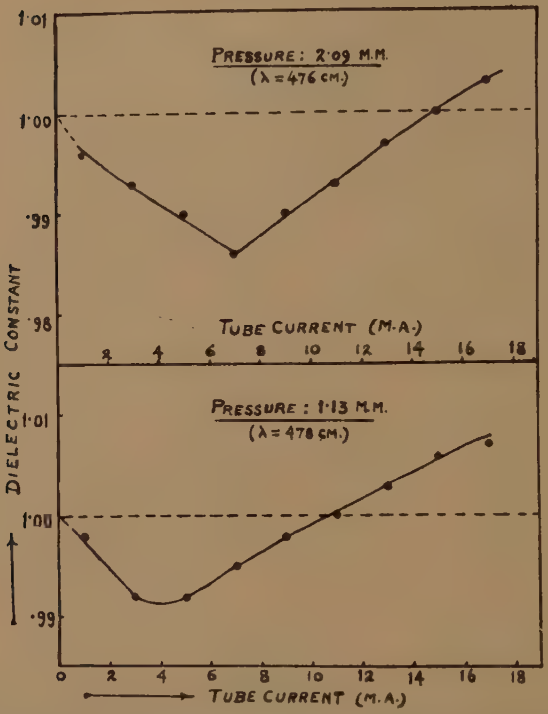


A typical semi-log curve.

3. *Experimental Results and their Interpretation.*

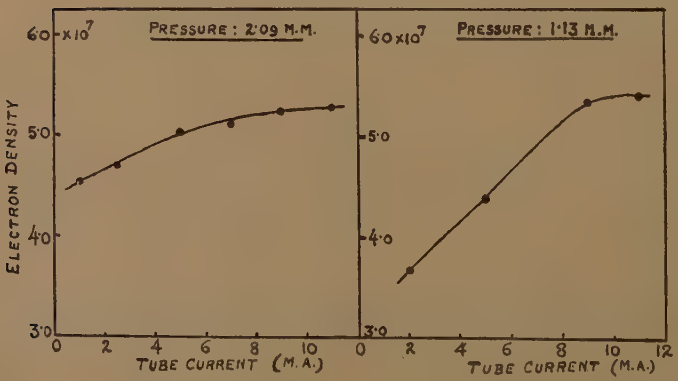
The variations of the effective dielectric constant of the ionized air for different values of the tube current for two sets of experiments with two different pressures in the discharge-tube are illustrated in fig. 4. In each case the value of the dielectric constant, which was less than unity, decreased with the increase of tube current up to a certain value of the latter. Beyond this value the dielectric constant gradually attained the value of unity and subsequently became greater than unity for larger tube currents. In fig. 5 are shown the values of the electron density of the core of the discharge in the region of the experimental condenser obtained by Langmuir's probe method for different tube currents for the same two pressures of the discharge-tube. In both instances the electron density was found to increase with the increase of the tube current. It should be noted that for 2·09 mm. pressure the value of the dielectric constant became unity and tended to exceed unity

Fig. 4.



Variation of dielectric constant with tube current.

Fig. 5.



at 15 m.a., whereas for 1·13 mm. pressure the dielectric constant tended towards a value greater than one when the tube current was 11 m.a.

The values of the electron density N and kT were determined by Langmuir's method for 11 m.a. tube current for both the pressures; these values are shown in Table I.

The anomalous increase of the dielectric constant of the ionized air has already been explained as due to the effect of the positive ionic sheath formed on the inner surface of the discharge-tube. It is, however, known that as the current through the tube is increased the thickness of the sheath becomes smaller and smaller. It is therefore evident that the effective capacity of the experimental condenser would increase or tend to increase, although there is an increase of electron concentration with the increase of the tube current. Thus the effect of the ionic sheath is to increase the value of the effective dielectric constant of the ionized air. For smaller tube currents the sheath effect is negligible, so that the dielectric constant would steadily decrease with the increase of the tube current, and subsequently for larger tube currents the sheath effect would cause a deviation from the normal variation of the dielectric constant. It is therefore expected that for a certain value of the tube current the thickness of the sheath would be such as to make the effective dielectric constant of the ionized air equal to unity. Beyond this value of the current the value of the dielectric constant would, however, exceed unity as the thickness of the sheath becomes smaller.

TABLE I.

Tube Current : 11 m.a.

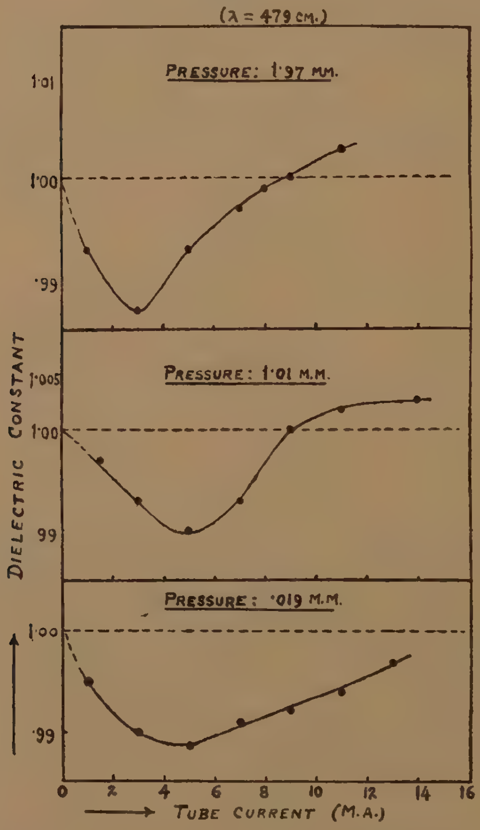
Pressure mm.	N per c.c.	kT.
2·09	$5{\cdot}3\times10^{7}$	$1{\cdot}45\times10^{-10}$
1·13	$5{\cdot}4\times10^{7}$	$1{\cdot}78\times10^{-10}$

The thickness of the ionic sheath is governed by the space-charge equation

$$i_{+}=\frac{A.V^{3/2}}{x^{2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where V is the fall of potential across the sheath, x the thickness of the sheath, i_{+} the positive ion current, and A a constant. Since the potential fall across the sheath ultimately depends on the velocity of the electrons relative to that of the positive ions, the sheath would assume the critical thickness for ϵ to become unity at a definite value of the electron velocity (or of kT), provided the positive ion current would remain the same. Ordinarily for very low pressures in the discharge-tube the positive ion current is small, and it increases with the increase of pressure attaining a saturation value * for the higher pressures. We have reasons to suppose that i_+ is more or less the same for the two tube pressures 2·09 mm. and 1·13 mm. in our experiments, so that it is expected that the value of the dielectric constant of the ionized air would become unity at the same kT-value of the electrons. In our experiments it was found that

Fig. 6.



for 2·09 mm. pressure the value of the dielectric constant tended to exceed unity when the tube current was 15 m.a., and for the lower pressure,

---

* The experiments of Gladys A. Anslow (Phys. Rev. xxv. p. 497 (1925)) showed that the number of positive ions per electron produced by fast moving electrons in air at low pressure increased rapidly with the increase of pressure when the pressure was very low and attained a saturation valve at some higher pressure. For 700 volts electrons the saturation was observed at about ·2 mm. of mercury.

1·13 mm., the corresponding value of the tube current was 11 m.a. We have also seen that the kT-values at 11 m.a. for the higher and lower pressures were respectively $1{\cdot}45\times10^{-10}$ and $1{\cdot}78\times10^{-10}$, as given in Table I. It can, however, be expected that the lower kT-value for the higher pressure at 11 m.a. would approach the higher value at 15 m.a. Hence it can be said that these experimental results are consistent with the theory of the ionic sheath.

For a more definite test, however, further experiments were performed. The same discharge-tube, which was of 2·5 cm. in diameter, was fitted up with larger aluminium electrodes exactly fitting into the tube. The measurements of the dielectric constant of the ionized air were carried out for three different pressures in the discharge-tube, viz., 1·97 mm., 1·01 mm., and ·019 mm. of mercury. The results are illustrated in fig. 6. It will be seen that for the two higher pressures the dielectric constant of the ionized air assumed the value of unity for the same value of the tube current, viz., 9 m.a., whereas when the pressure was reduced to about a hundredth of the highest pressure, the value of the dielectric constant was still less than unity even when the tube current was 13 m.a. The values of the electron density N and kT by the probe method were determined for 9 m.a. tube current in the three cases. These are given in Table II.

TABLE II.

Tube Current : 9 m.a.

Pressure (in mm.)	N per c.c.	kT.
1·97	$3{\cdot}15\times10^{7}$	$\cdot85\times10^{-10}$
1·01	$3{\cdot}45\times10^{7}$	82×10^{-10}
·019	$3{\cdot}6\ \times10^{7}$	$1{\cdot}31\times10^{-10}$

It is evident from the table that the kT-values at 9 m.a. tube current were practically the same for the two higher pressures. For the higher pressures the positive ion currents at this value of the tube current could be regarded as more or less equal, so that the dielectric constant of the ionized air would attain the value of unity at the same value of the tube current, viz., 9 m.a., where the kT-values were equal. In the case of the lowest pressure (·019 mm.) the kT-value of the electrons at 9 m.a. was as high as $1{\cdot}31\times10^{-10}$, but for such low pressures the positive ion current must needs be considerably less than that at the higher pressures, so that, according to the space charge equation (7), the thickness of the ionic sheath would be expected to be much larger. This would explain why the requisite thickness required for ϵ to become unity was not attained even for a tube current of 13 m.a.

4. *Summary.*

In the present investigation experiments were performed with external condenser plates round a discharge-tube on the effective dielectric constant of ionized air inside such a tube having different pressures and carrying different currents. For each different tube current and pressure the electron density and the electron temperature in the region of the experimental condenser were determined by the probe method of Langmuir. The results of the experiments, which revealed the previously reported anomalous variation of the dielectric constant of the ionized air, were found consistent with the theory that the anomalous values are due to the formation of a positve ionic sheath on the inner surface of the discharge-tube.

Physics Department,
Dacca University.

XXXVI. *The Dielectric Constant and Electrical Conductivity of Gases and Vapours Ionized by X-rays at Ultra-high Radio Frequency.*

By S. R. Khastgir, D.Sc.(Edin.), Mem.I.R.E., and Shah Md. Fazlur Rahman, M.Sc., Dacca University *.

[Received March 31, 1939.]

1. *Introduction.*

Some years ago Barton and Kilby † made an interesting investigation on gaseous ionization produced at atmospheric pressure by X-rays or radium bromide. Their experimental apparatus included an oscillatory circuit in which there was a condenser of two plates between which the ionization was produced. A change in the natural frequency of the oscillatory circuit was detected by a suitable modification of the Fleming cymometer indicating an apparent change in the capacity of the experimental condenser which was attributed to the effect of the conductance of the gas acting as a leak in parallel with the condenser. According to Appleton and Childs ‡, the work of Barton and Kilby required a different interpretation and probably did show a reduction of dielectric constant below unity. In view of the growing interest in the subject of radio-frequency properties of ionized media, it was considered desirable to undertake a careful investigation on the subject, using different gases and vapours ionized by X-rays. A report of the results of this investigation will be given in this paper. Four different gases (or vapours) were employed, viz.:—

(1) Air.
(2) Sulphur dioxide.
(3) Argon.
(4) A mixture of methyl iodide and ethyl bromide.

Two distinct sets of experiments were performed with each gas or vapour, employing oscillations of wave-lengths of 460–480 cm. :—

(1) Study of the dielectric constant change and the conductivity effect in the different gases ionized by varying intensities of X-rays

---

* Communicated by the Authors.
† Barton and Kilby, Phil. Mag. xxvi. ser. 6, (1913).
‡ Appleton and Childs, Phil. Mag. x. p. 969 (1930).

obtained by filtration through aluminium sheets, the pressure of each gas remaining the same.

(2) Study of the dielectric constant change and the conductivity effect in the gases ionized by X-rays under varying pressures (the maximum range being from about 10 cm. to about 130 cm. of mercury).

A Coolidge tube was employed for the production of X-rays required for gaseous ionization.

2. *Theoretical Considerations.*

(a) *Consideration of the Dielectric Constant Change and the Conductivity Effect.*

Let C_0 be the capacity of a condenser filled with some gas and fixed at the end of a horizontal pair of Lecher wires. When the ultra-high frequency oscillations from an oscillator are induced into the Lecher system, the voltage or the current resonance adjustments can be made by sliding a copper bridge short-circuiting the two wires. When the gas inside the condenser is ionized by X-rays there is usually a shift of the resonance point. For voltage resonance, *i. e.*, for the maximum voltage across the experimental condenser, the shift is due to the change of the dielectric constant of the gas on being ionized. For current resonance, however, the shift is due to the combined effect of the dielectric constant change and of the conductivity acquired by the gas on ionization.

Let ρ be the leakage resistance across the condenser when the gas inside it is ionized and ϵ the dielectric constant of the ionized gas. If w be the angular frequency of the field, the vector impedance z_1' across the "leaky" condenser is given by

$$\frac{1}{z_1'}=\frac{1}{\rho}+j\omega(\epsilon.C_0),$$

so that
$$z_1'=\frac{\rho}{1+\rho^2\omega^2(\epsilon.C_0)^2}-\frac{j\omega(\epsilon.C_0)\rho^2}{1+\rho^2\omega^2(\epsilon.C_0)^2}.$$

If the resonance length l is changed to l', on the gas being ionized, the total impedance of the Lecher circuit for wave-length λ is given by

$$z_2'=j\sqrt{\frac{L}{C}}\tan\frac{2\pi l'}{\lambda}+\frac{\rho}{1+\rho^2\omega^2(\epsilon.C_0)^2}-\frac{j\rho^2\omega(\epsilon.C_0)}{1+\rho^2\omega^2(\epsilon.C_0)^2},$$

where L and C are the inductance and capacity per unit length of the Lecher wires.

For current resonance the reactance part must be zero, that is,

$$j\sqrt{\frac{L}{C}}\tan\frac{2\pi l'}{\lambda}-\frac{j\omega\rho^2(\epsilon.C_0)}{1+\rho^2\omega^2(\epsilon.C_0)^2}=0,$$

or

$$\sqrt{\frac{L}{C}}\tan\frac{2\pi l'}{\lambda}=\frac{1}{\omega(\epsilon.C_0)+\dfrac{1}{\rho^2\omega(\epsilon.C_0)}} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Considering the conductivity term

$$\sigma=\frac{1}{4\pi\rho C_0}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

we can write

$$\sqrt{\frac{L}{C}}\tan\frac{2\pi l'}{\lambda}=\frac{1}{\omega(\epsilon.C_0)\left[1+\left(\dfrac{4\pi\sigma}{\epsilon\omega}\right)^2\right]}. \quad . \quad . \quad . \quad . \quad (3)$$

When the gas inside the condenser is not ionized $\sigma=0$, $l=l'$, and $\epsilon \gtrsim 1$. Hence (3) is reduced to

$$\sqrt{\frac{L}{C}}\tan\frac{2\pi l}{\lambda}=\frac{1}{\omega C_0}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

From (3) and (4)

$$\frac{\tan\dfrac{2\pi l}{\lambda}}{\tan\dfrac{2\pi l'}{\lambda}}=\epsilon\left[1+\left(\frac{4\pi\sigma}{\epsilon\omega}\right)^2\right] \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

If $\sigma=0$, $\dfrac{\tan\dfrac{2\pi l}{\lambda}}{\tan\dfrac{2\pi l'}{\lambda}}=\epsilon$, and for $\epsilon<1$ the shift $(l'-l)>0$ or $l'>l$.

For a finite value of the conductivity σ, it can be seen from (5) that for $\epsilon<1$, the shift $(l'-l) \gtrless 0$, according as

$$\left(\frac{4\pi\sigma}{\omega}\right)^2 \lessgtr \epsilon(1-\epsilon).$$

Thus for a large value of the conductivity, when the length adjustments for the current resonance are made, it is possible to have a negative value of the shift $(l'-l)$, *i. e.*, $l'<l$ even when $\epsilon<1$. It can also be seen that the shift of the voltage resonance point is larger than the corresponding shift of the current resonance point.

(b) *Theory of the Method of Measuring the Electrical Conductivity of the Gaseous Ionization.*

The measurement of the high-frequency electrical conductivity of the ionized gas is made by measuring the change of attenuation constant of the Lecher system when the gas inside the experimental condenser is ionized by X-rays. The change in the attenuation of the Lecher system can be obtained from the change in the " half-width " of the resonance curves constructed with X-rays off and on. In the resonance curves the deflexions of the galvanometer in the anode circuit of the valve-detector used for determining the resonance points are to be plotted against the scale divisions of the Lecher wires. By " half-width " is meant the distance between the points on either side of the resonance point when the galvanometer deflexion becomes half that at resonance.

When the gas inside the experimental condenser is not ionized, assuming no natural leakage of the condenser, the attenuation constant of the Lecher system is approximately given by

$$\alpha_0=\frac{R}{2}\sqrt{\frac{C}{L}}, \qquad (6)$$

where R, L, and C are the resistance, inductance and capacity per unit length of the Lecher system. When R is extremely small, the surge-impedance of the Lecher system is given by

$$Z_0=\sqrt{\frac{L}{C}},$$

so that

$$\alpha_0=\frac{R}{2Z_0}. \qquad (7)$$

When the gas is ionized, there will be a leakage of resistance ρ across the experimental condenser. This parallel resistance can be replaced by a series resistance $(r.l')$, where r=resistance per unit length distributed uniformly over the entire resonance length l'. The attenuation constant under this condition is given by

$$\alpha=\frac{R+r}{2Z_0}. \qquad (8)$$

Thus from (7) and (8)

$$r=2Z_0(\alpha-\alpha_0),$$

or

$$(r.l')=2Z_0(\alpha-\alpha_0)l' \qquad (9)$$

The dielectric constant $\epsilon\simeq 1$, so that the leakage resistance is given by

$$\rho=\frac{1}{\omega^2 C_0^{\,2}(r.l')} \qquad (10)$$

C_0 can be eliminated from (10) by making use of the current resonance condition when the gas inside the experimental condenser is not ionized, viz. :—

$$\sqrt{\frac{L}{C}}\tan\frac{2\pi l}{\lambda}=\frac{1}{\omega C_0},$$

or
$$C_0=\frac{1}{\omega Z_0\tan\frac{2\pi l}{\lambda}}.$$

Substituting this value C_0, we obtain from (9) and (10)

$$\rho=\frac{Z_0.\tan^2\frac{2\pi l}{\lambda}}{2(\alpha-\alpha_0)l'}\quad\text{. . . . .}\quad(11)$$

Again, remembering $\rho=\frac{1}{4\pi\sigma C_0}$ and substituting the value of C_0, we get

$$\rho=\frac{Z_0.\omega.\tan\frac{2\pi l}{\lambda}}{4\pi\sigma}\quad\text{. . . . . . .}\quad(12)$$

From (11) and (12)

$$\left.\begin{aligned}\sigma&=\frac{\omega(\alpha-\alpha_0)l'}{2\pi\tan\frac{2\pi l}{\lambda}}\\&=\frac{c(\alpha-\alpha_0)l'}{\lambda\tan\frac{2\pi l}{\lambda}}\end{aligned}\right\}\quad\text{. . . . .}\quad(13)$$

Since the attenuation constants α_0 and α are given by

$$\alpha_0=\frac{2\pi e_0}{n\lambda^2},$$

$$\alpha=\frac{2\pi e}{n\lambda^2},$$

where e_0 and e are the " half-widths " of the resonance curves constructed for the non-ionized and the ionized gas respectively, and n is the order of the potential node, the relation (13) becomes

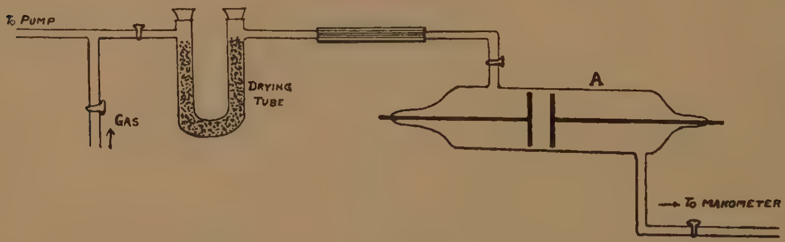
$$\sigma=\frac{2\pi c(e-e_0)l'}{n\lambda^3\tan\frac{2\pi l}{\lambda}}\quad\text{. . . . .}\quad(14)$$

The value of the conductivity σ can be calculated from (14) by measuring the difference of the half-widths and the current resonance lengths l and l' for the non-ionized and the ionized gas respectively.

3. *Experimental Details.*

Two parallel circular disks of diameter 2·3 cm., placed vertically at a distance of ·5 cm. from each other, were fixed inside a suitable glass cylinder, so as to form the terminal condenser A of a horizontal pair of Lecher wires consisting of No. 18 bare copper wire. The cylinder had an inlet and an outlet glass tube provided with stopcocks. The outlet glass tube was led on to one limb of a manometer, the other limb opening out into the air. There was provision for introducing through the inlet tube the gas under observation into the cylinder after having dried it by passing through a calcium chloride tube. (See fig. 1.) A Cenco hyvac pump was employed for initial evacuation of the glass cylinder. This glass cylinder with the aluminium disks was placed immediately in front of a circular aperture of 1″ diameter on the side of a box lined with thick lead sheets. A Coolidge X-ray tube was placed inside the box with its tungsten anticathode at a suitable distance from, and in line with, the terminal parallel plate condenser of the Lecher system. The ultra-high frequency oscillator was mounted on a suitable stand in such a way that all its components were above the Lecher wires with only the inductance loop L of the oscillatory circuit hanging vertically down directly in front of the terminal condenser. The ultra-high frequency oscillations were thus induced into the Lecher system. Immediately beyond the inductance loop a lead screen *s* with two holes for the Lecher wires to pass through was held, thus preventing the ionization of air behind this screen by the X-rays after having passed through the experimental condenser. The diagram of the whole arrangement is shown in fig. 2 (*a*).

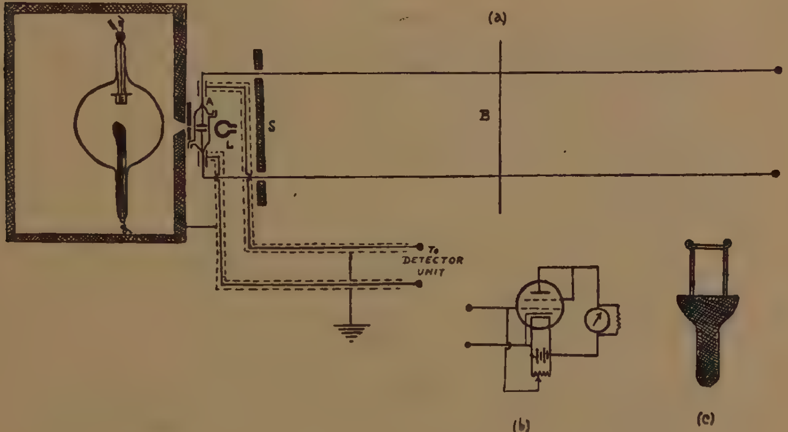
Fig. 1.



Adjustments of Lecher wire lengths for voltage resonance, *i. e.*, for maximum voltage across the plates of the experimental condenser or for current resonance were made by sliding a short-circuiting copper bridge B along the wires. In the case of voltage resonance adjustments, the detector-unit, which was a properly shielded valve-voltmeter comprising a detector valve with a mirror galvanometer in the anode circuit (fig. 2 (*b*)), was coupled to the Lecher wire system *at the condenser end* in the usual way, *i. e.*, by means of two small condensers consisting of short brass tubes insulated from the Lecher wires by glass sleeves. For current resonance, on the other hand, the input terminals of the detecting unit were disconnected from the input end of the Lecher wire system and directly joined to the two ends of the short-circuiting bridge. When the current through the short-circuiting bridge was a maximum, the extremely small voltage across it must also be maximum and would be indicated by the sensitive valve-detector. The short-circuiting bridge is shown in fig. 2 (*c*). It consisted of No. 18 bare copper wire bent into a rectangular frame and mounted on a suitable ebonite handle. The side of the frame which was used as the bridge was filed so as to have a prismatic shape, so that the sharp edge of the prism touched the wires when the resonance adjustments were made.

The Coolidge tube was worked with a 12″ spark induction coil and a mercury interrupter. The filament was fed by accumulators of a requisite voltage. The tube current, as measured by a milliammeter in the tube circuit, was kept at a fixed value during a set of observations ; the filament current was also kept constant. For the “ off ” measurements, however, the X-ray tube was kept running, but the X-rays were cut off by inserting a lead sheet in the aperture in front of the experimental condenser. The lead sheets covering the sides of the wooden box enclosing the X-ray tube were all earthed.

Fig. 2.

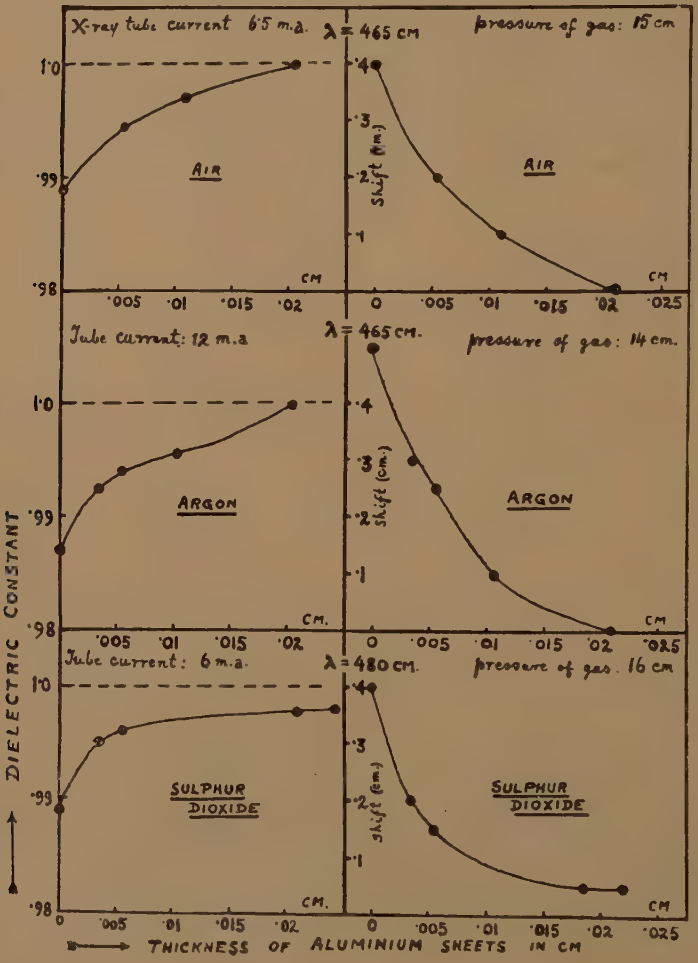


4. *Change in the Dielectric Constant of the Different Gases and Vapours when Ionized by X-rays of Varying Intensities at a Definite Pressure.*

(a) *Experimental Results.*

When the length adjustments for maximum voltage across the plates of the experimental condenser containing gas or vapour were made, it was invariably found that when the gas was ionized by X-rays the resonance l' was *always* greater than the corresponding length l when the gas was not ionized, showing that the value of the effective dielectric constant was reduced in the ionized condition. Varying intensities of X-rays obtained by passing through varying thickness of aluminium filters produced different amounts of ionization, and the shifts $(l'-l)$ were found to increase gradually with increasing amounts of ionization. The results are illustrated in fig. 3 for air, sulphur dioxide, and argon. In the diagram for each gas, the shift and the corresponding calculated value of the

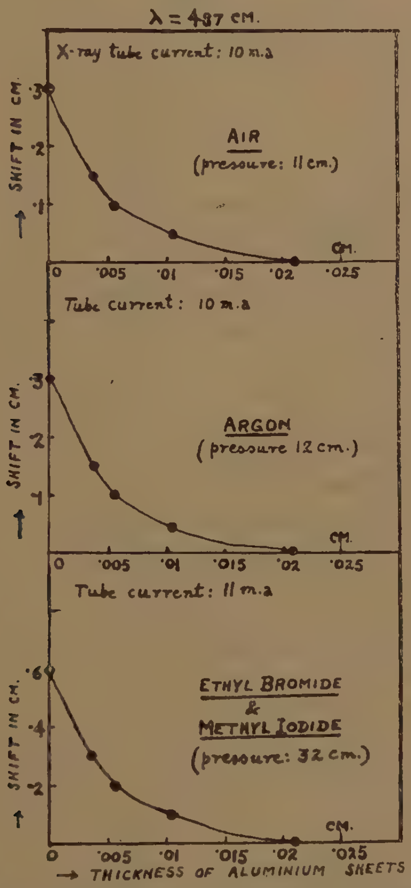
Fig. 3.



dielectric constant of the ionized gas are both given for increasing thicknesses of absorbing filters *.

Similar sets of experiments were performed where the length adjustments were made for current resonance. Here also the resonance length l'_i for the ionized gas was *always* greater than the corresponding length l^i when the gas was not ionized. The shift is to be attributed to combination of the two effects : (1) the effect of the dielectric constant change, and (2) the conductivity effect. The conductivity was, however, found to be extremely small (the results of the conductivity measurements are given in a subsequent section), so that the observed shift indicated mainly the change in the dielectric constant when the gas was ionized.

Fig. 4.



---

* The dielectric constant of each gas in the non-ionized condition was assumed unity.

The experimental results for air, argon, and a mixture of ethyl bromide and methyl iodide are illustrated in fig. 4. Barton and Kilby, however, obtained positive as well as negative values of the shift of the resonance point for ionized air.

(b) *Conclusions from the Experimental Results.*

Since the collisional frequency in the ionized gas or vapour in these experiments is very much higher than the angular frequency of the field, the change in the dielectric constant of the gas (or vapour) on ionization is given by

$$\Delta\epsilon = -\frac{4\pi N e^2}{m\nu^2},$$

where N is the ionic density, ν the collisional frequency, and e and m are the charge and mass of an ion. The ions are of molecular dimensions, and since under the conditions of our experiments the ionic density cannot be very large, and since the collisional frequency was of the order of 10^{12}, a perceptible change of dielectric constant was not theoretically possible. And yet when the length adjustments were made for voltage resonance, the observed positive value of the shift of the resonance point strongly indicated a reduction in the dielectric constant. We have therefore to postulate that in the process of gaseous ionization by X-rays or γ-rays, each of the photo-electrons ejected remains as a free electron for a small fraction of the time, after which it attaches itself to a molecule or a cluster of molecules and becomes an ion. Owing to the exceedingly small value of the mass of an electron, a very short period of freedom can imply a perceptible change in the dielectric constant. This postulate of the free state of the photo-electrons for a short time after their ejection has probably its support from the marked difference that is observed between the mobilities of the positive and negative ions in the simple gases *.

5. *Change in the Dielectric Constant of the Different Gases and Vapours Ionized by X-rays for Varying Pressures, the Intensity of X-rays remaining Constant.*

(a) *Experimental Results.*

The gas or vapour under test was kept at different pressures, and the X-ray tube current and the filament current were maintained at steady values. The shift $(l'-l)$ of the resonance point both for voltage and current resonance on ionizing the gas was found to be *always* positive and *independent of the pressure of the gas.* In the case of air, argon, and sulphur dioxide the range of pressure was from about 10 cm. to about 130 cm. of mercury, and for the mixture of ethyl bromide and methyl iodide the range was from 15 to 76 cm. The shift was, however, found to be of different magnitude for the different gases. It should be mentioned here that the shift of the current resonance point on the gas being ionized by X-rays was of the same order of magnitude as the corresponding shift observed in the case of voltage resonance. This was expected, since the conductivity of the ionized gas was only of the order of 10^4 or 10^5 e.s.u.

---

* A strong argument against this view is that the ratio of the mobilities is much the same for gases which, according to Loeb's results (Phys. Rev. xxvii. p. 811 (1926)), show great differences in the ease with which they combine with electrons. The difference in the mobilities of the positive and the negative ions can, however, be explained in by the "cluster" theory. (*Vide* J. J. Thomson's 'Conduction of Electricity through Gases,' vol. i. p. 180 (1928).)

(b) *Discussion of the Results.*

It can be shown that ordinarily the ionic density in the case of gaseous ionization remains practically constant for different pressures, except when the pressure is very high. When the gas is ionized, the collisions between positive and negative ions result in a certain amount of recombination. After an interval of time the gas approximates to a steady state. If n_0 be the ionic density (*i. e.*, number of ions per c.c.) in the steady state q the number of ions produced per c.c. per sec. by the ionizing agent, and α the coefficient of recombination, then

$$n_0=\sqrt{\frac{q}{\alpha}}.$$

When the ionization is produced by X-rays, q is proportional to the pressure, and it is also known from the results of several experimenters that when the pressure is not too high, the coefficient of recombination α is proportional to the pressure. Thus when q and α are both proportional to the pressure, the density of the ions n_0 in the steady state must be independent of the pressure.

It should be remembered here that if the shift of the voltage resonance point when the gas is ionized by X-rays is to be attributed to the change in the dielectric constant, we shall be concerned with the number of electrons per c.c. which remain in the free state for a small part of the time of exposure of X-rays before these electrons attach themselves to molecules to form ions of molecular dimensions. Accepting our experimental results, we can therefore conclude that in the Eccles-Larmor expression

$$\Delta\epsilon=-\frac{4\pi Ne^2}{mv^2},$$

$\left(\frac{N}{\nu^2}\right)$ ought to be independent of the pressure.

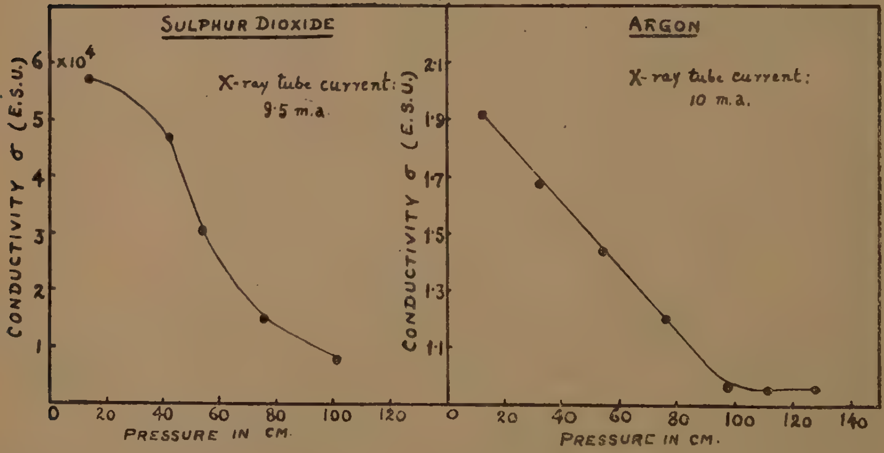
Here N is the number of electrons per c.c. which remain free for a part of the time of exposure of X-rays and ν the electron collision frequency. Since ν is directly proportional to the pressure, the constancy of the dielectric constant change on ionizing the gas or vapour for different pressures can be explained if $N \propto p^2$.

6. *Persistence of the Effect of X-rays.*

In most cases it was found that the increased resonance length of the Lecher wires observed when the gas inside the terminal condenser was ionized by X-rays was not restored to the original length when the X-rays were blocked out by a lead screen. After waiting for some length of time, however, the resonance length tended to approach its original smaller value corresponding to the non-ionized condition of the gas. For this reason for each observation fresh gas was introduced into the experimental condenser after evacuating it more than once very carefully.

Fig. 5.

λ = 481 cm.



7. *Measurement of the H.F. Electrical Conductivity of the Different Gases Ionized by X-rays.*

The theory of the method adopted for determining the value of the conductivity of the ionized medium is already given. The experimental results for sulphur dioxide and argon are illustrated in fig. 5. In the case of sulphur dioxide the observed maximum conductivity was $5{\cdot}7 \times 10^4$ e.s.u., and in the case of argon it was $1{\cdot}9 \times 10^5$ e.s.u. The wave-length of the alternating field was 481 cm. The resonance curves were drawn for the second potential node on the Lecher wires.

Under the conditions of our experiments, the conductivity of a gas is given by

$$\sigma = \frac{Ne^2}{mv},$$

where N is the number of gaseous ions per c.c. which in the steady state is independent of the pressure. If we now suppose that the collision frequency is directly proportional to the pressure, it follows that the conductivity should be inversely proportional to the pressure. This is supported by our experimental results, which clearly indicated that the conductivity decreased as the pressure was increased within the range of our observations.

8. *Summary.*

In the present investigation Barton and Kilby's experiments were performed with a different arrangement of apparatus. The terminal condenser of a horizontal pair of Lecher wires into which ultra-high frequency oscillations were induced was filled with gas or vapour at a definite pressure. Length adjustments on the Lecher wires for maximum voltage across the terminal condenser and also for current resonance were made by sliding a short-circuiting bridge across the wires—suitable arrangements having been made with a valve-detecting unit to indicate the resonance point. The gas or vapour under test was ionized by X-rays from a Coolidge tube. The shift of the voltage or the current resonance point on ionizing the gas or vapour was then studied for four different gases and vapours, viz., air, sulphur dioxide, argon and a mixture of methyl iodide and ethyl bromide. In all the cases a *positive* shift was observed. The shift was found gradually to diminish with the reduction in the X-ray intensity. Again on varying the pressure of the gas or vapour and keeping the intensity of the X-rays fixed, the shift was found independent of the pressure. These results indicated that there must be a diminution in the effective dielectric constant of the gaseous medium on ionization. This change in the dielectric constant is possible if it is assumed that the ejected photo-electrons remain in a free state at least for a small part of the time of exposure to X-rays.

The high-frequency electrical conductivity of argon and sulphur dioxide ionized by X-rays was also measured for different pressures by a resonance curve method. The conductivity was found to be of the order of 10^4 or 10^5 e.s.u. and was found to diminish with the increase of pressure, as was expected from theory. Since the conductivity was very small, the voltage resonance condition was sensibly the same as the current resonance condition.

Once the gas or vapour was exposed to X-rays it was observed that in most cases the resonance length did not revert to the smaller value when the X-rays were blocked by a lead shutter. The effect of the X-rays was found to persist for a long time. The shift of the resonance point on ionization was therefore measured each time with fresh gas inside the experimental condenser.

Physics Department,
Dacca University.

XXXVII. *The Temperature Variation of Viscosity in Non-Polar and Dipolar Inorganic Liquids.*

By S. T. Bowden, D.Sc., and A. R. Morgan, M.Sc.,
University College, Cardiff *.

[Received July 26, 1939.]

Introduction.

The change of viscosity with temperature has been extensively studied in the case of a large number of liquids, and many equations have been proposed to represent the observed decrease of viscosity with rising temperature. The majority of these equations were essentially empirical in character and proved to be of limited applicability. During the last twenty years much attention has been given to the exponential relation, which may be expressed in its simplest form as

$$\eta = Ae^{\frac{B}{T}}, \quad \ldots \ldots \ldots \quad (1)$$

where η is the viscosity, T the absolute temperature, and A and B are constants. The first derivation of the exponential formula appears to have been made in 1912 by Arrhenius ⁽¹⁾, although the first publication is due to de Guzman ⁽²⁾. The relation has been the subject of further study by a number of workers ⁽³⁾ and has been found to be remarkably successful in representing the change of viscosity of a variety of liquids over a wide range of temperature.

The simple formula was derived by Andrade ⁽⁴⁾ on the assumption that at extreme libration the molecules of one layer come into contact with the molecules of another layer and enter into a brief temporary union, so that they attain a common velocity of translation. The mechanism underlying the transfer of momentum from layer to layer thus differs radically from that assumed in Maxwell's theory of gaseous viscosity, which depends upon actual interchange of molecules between the layers. In a liquid the probability of union will fall with the increase in thermal agitation resulting from a rise of temperature. The simple exponential formula may be derived on these premises, and by taking into account the influence of thermal expansion may be further modified to

$$\eta v^{\frac{1}{3}} = Ae^{\frac{c}{vT}},$$

---

* Communicated by the Authors.

where v is the specific volume, and A and c are constants. This relation represents the temperature variation of viscosity with considerable accuracy in the case of a large number of organic liquids, but, apart from the halogens, water, and a few metals and salts, it has not been applied to inorganic liquids.

Object of the Investigation.

For the purpose of ascertaining the behaviour of non-polar and dipolar inorganic liquids in this connexion, we have measured the viscosity of silicon tetrachloride, phosphorus trichloride, arsenic trichloride, sulphur monochloride, phosphorus oxychloride, thionyl chloride, and sulphuryl chloride, and have further tested the applicability of the equation in the case of stannic chloride and bismuth trichloride. As may be seen from the accepted electronic formulations, some of the non-polar and dipolar substances simulate the structures of organic compounds, while phosphorus oxychloride, thionyl chloride and sulphuryl chloride have co-ordinate linkages, and thus differ fundamentally from the ionic and metallic liquids which have been studied previously.

TABLE I.

	B.-p.	Pressure.	$d^{25°}_{4°}$.
	°	mm.	
Silicon tetrachloride	57·4	772	1·4707
Phosphorus trichloride	75·7	758	1·5659
Arsenic trichloride	129·0	771	2·1497
Sulphur monochloride	42·0	23	1·6714
Phosphorus oxychloride ...	105·9	762	1·6676
Thionyl chloride..........	76·5	766	1·6290
Sulphuryl chloride	69·2	754	1·6570

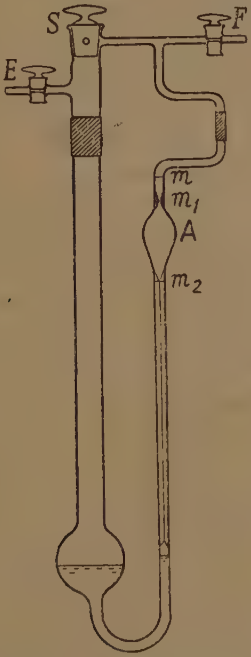
Experimental.

Purification of Materials.—The inorganic liquids were allowed to stand for a short time over activated silica gel for the purpose of adsorbing hydrogen chloride, and were then subjected to three separate fractionations in an all-glass apparatus incorporating a long Dufton column fitted with a ground-glass spiral. The intrusion of moisture during these operations was prevented by guard-tubes of calcium chloride and silica gel. The values of the boiling-point and density of the pure liquids are given in Table I.

The boiling-point of arsenic trichloride differs by 7° from that given in the 'International Critical Tables,' but is in close agreement with the values recorded in Landolt-Bornstein Tabellen. Sulphur monochloride suffers decomposition when heated to the boiling-point under ordinary pressure, and the material was therefore subjected to fractionation under reduced pressure. The liquid was free from the red colour associated with the presence of the higher chlorides of sulphur.

The organic liquids employed for standardizing the viscometers were purified by the following methods. " Analar " benzene, extra pure for molecular weight determinations, was shaken with concentrated sulphuric acid, washed with water, dried over calcium chloride, and finally distilled from phosphorus pentoxide (free from trioxide) ; b.-p. 80·0°/762 mm.,

Fig. 1.



Viscometer for reactive liquids.

$d^{25^\circ}_{4^\circ}$ 0·8732. " Analar " bromobenzene was dried over calcium chloride and thrice fractionated in a Dufton apparatus ; b.-p. 154°/762 mm., $d^{80^\circ}_{4^\circ}$ 1·4130. " Analar " ether was dried over freshly drawn sodium wire and then distilled over phosphorus pentoxide ; b.-p. 34·6°/765 mm., $d^{0^\circ}_{4^\circ}$ 0·7360.

Viscosity Measurements.—Viscometers of the type shown in fig. 1 were constructed according to Specification No. 188 of the British Engineering Standards Association, and the design embodied certain features recommended in recent practical treatises (5). The capillary tubes, specially selected with respect to uniformity of bore, were 7–10 cm. in length and 0·36–0·50 mm. in diameter according to the viscosity of the liquid under investigation, and the upper bulb A was of pear-shape to facilitate draining of the liquid. Since the liquids under investigation are very hygroscopic and easily hydrolysed, the limbs of the viscometer were provided with a glass connector, so that the measurements could be carried out in a closed system to obviate the intrusion of moisture. This connector with the special stopper S was found to be superior to the types previously described, since it enables the viscometer to be charged with liquid while the connector is actually in position. Guard-tubes containing anhydrous calcium chloride and plugs of cotton-wool to serve as dust-traps were attached to E and F. With the viscometer rigidly fixed on a brass frame, the liquid was introduced through S by means of a long standard pipette fitted with a stopcock and guard-tube of calcium chloride. The viscometer was placed in the thermostat and set vertically to a plumb-line, and allowed to attain the temperature of the bath. With the stopper S arranged as in fig. 1 a source of dry, compressed air was connected to the side-tube E, so that the liquid was forced up to the initial reference mark m. The compressed air unit consisted of a Winchester bottle charged with air dried *in situ* by phosphorus pentoxide; it was fitted with a needle-valve which provided a delicate means of controlling the rate of rise of liquid through the capillary. The stopcocks E and F were closed at this stage and the compressed air unit removed. The stopper S (with a 3 mm. orifice) was then turned to bring the two limbs into communication, and the time of flow between m_1 and m_2 noted by means of a stop-watch previously calibrated against a standard chronometer. The use of compressed air in this way preserved a positive pressure within the apparatus and thus prevented the intrusion of moisture.

The thermostats were electrically controlled to 0·02° and all thermometers were standardized against instruments calibrated at the National Physical Laboratory. For experiments at the higher temperatures, previous workers have pointed out the necessity of applying a correction to compensate for the expansion of the liquid and the resultant change in pressure head. For the purpose of the present experiments at 0–35°, the liquids were measured at room-temperature, but for those at 50°, 65°, and 80° the pipette and liquids to be measured were kept at the temperature of the bath. The kinetic energy correction was made by employing the equation

$$\eta=d\left(\mathrm{A}t-\frac{\mathrm{B}}{t}\right),$$

where d is the density, t the flow-time, and A and B are constants which were evaluated by measuring the times of flow of two liquids of different kinematic viscosity. At 0° the liquids employed for the calibration were water and ethyl ether, for the range 15–50° water and benzene, and for the range 65–80° water and bromobenzene.

The density determinations were made in a 25 c.c. bottle of thin glass and fitted with a specially ground glass stopper. In view of the extreme susceptibility of the liquids to suffer hydrolysis, the determinations were carried out as rapidly as was consistent with accuracy, and in the case of the more readily hydrolysable substances the adjustment of the liquid level was made while a stream of dry nitrogen passed upwards through the neck of the density bottle. The weights employed were calibrated by Richards's method, and the usual corrections were applied for buoyancy.

Results.

The results of the density and viscosity determinations are recorded in Table II., where d is the density and η is the viscosity expressed in poises.

TABLE II.

Silicon Tetrachloride.

Temp. °C.	d.	η.
0	1·5211	0·00551
15	1·4914	0·00445
25	1·4707	0·00417
35	1·4500	0·00387

Phosphorus Trichloride.

Temp. °C.	d.	η.
0	1·6114	0·00656
15	1·5844	0·00560
25	1·5659	0·00501
35	1·5470	0·00479
50	1·5181	0·00453

Arsenic Trichloride.

Temp. °C.	d.	η.
0	2·2038	0·01531
15	2·1715	0·01222
25	2·1497	0·01068
35	2·1278	0·00977
50	2·0960	0·00847
65	2·0622	0·00745
80	2·0296	0·00648

Sulphur Monochloride.

Temp. °C.	d.	η.
0	1·7102	0·01236
15	1·6873	0·01037
25	1·6714	0·00925
35	1·6556	0·00855
50	1·6326	0·00750
65	1·6090	0·00660
80	1·5850	0·00585

Phosphorus Oxychloride.

Temp. °C.	d.	η.
0	1·7122	0·01445
15	1·6854	0·01141
25	1·6676	0·00999
35	1·6490	0·00887
50	1·6207	0·00737
65	1·5924	0·00646
80	1·5641	0·00554

Thionyl Choride.

Temp. °C.	d.	η.
0	1·6766	0·00814
15	1·6483	0·00680
25	1·6290	0·00603
35	1·6092	0·00577
50	1·5797	0·00534
65	1·5498	0·00505

TABLE II. (*cont.*).

Sulphuryl Chloride.

Temp. °C.	*d*.	η.
0	1·7091	0·00931
15	1·6778	0·00762
25	1·6570	0·00674
35	1·6360	0·00632
50	1·6044	0·00562

Discussion of Results.

The viscosities calculated from the Andrade formula are given in Table III., which also includes the values of the constants A and *c*, and estimates of the percentage deviation.

TABLE III.

Silicon Tetrachloride.

A=0·0009303. *c*=277.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
0	0·0051	0·00501	+9·1
15	0·00445	0·00446	−0·2
25	0·00417	0·00415	+0·5
35	0·00387	0·00388	−0·3

Stannic Chloride.

A=0·001212. *c*=220.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
25	0·00919	0·00814	+11·4
30	0·00806	0·00778	+ 3·5
40	0·00725	0·00724	+ 0·1
50	0·00668	0·00675	− 1·1
70	0·00600	0·00595	+ 0·8

Phosphorus Trichloride.

A=0·001830. *c*=163.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
0	0·00656	0·00562	+14·3
15	0·00560	0·00523	+ 6·6
25	0·00501	0·00500	+ 0·2
35	0·00479	0·00480	− 0·2
50	0·00453	0·00453	0

Arsenic Trichloride.

A=0·0008006. c=325.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
0	0·01531	0·01436	+6·2
15	0·01222	0·01202	+1·6
25	0·01068	0·01077	−0·9
35	0·00977	0·00972	+0·5
50	0·00847	0·00844	+0·4
65	0·00745	0·00740	+0·7
80	0·00648	0·00657	−1·1

Bismuth Trichloride.

A=0·007932. c=449.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
260	0·320	0·316	+1·3
270	0·295	0·293	+0·8
280	0·270	0·271	−0·4
290	0·250	0·251	−0·4
300	0·230	0·233	−1·3
310	0·215	0·218	−1·4
320	0·205	0·203	+1·0
330	0·190	0·190	0
340	0·180	0·178	+1·1

Sulphur Monochloride.

A=0·0008373. c=400.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
0	0·01236	0·01227	+0·7
15	0·01037	0·01038	0
25	0·00925	0·00937	−1·3
35	0·00855	0·00851	+0·5
50	0·00750	0·00745	+0·7
65	0·00660	0·00659	+0·1
80	0·00585	0·00588	−0·5

Phosphorus Oxychloride.

A=0·0005014. c=506.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
0	0·01445	0·01429	+1·1
15	0·01141	0·01149	−0·7
25	0·00999	0·01006	−0·7
35	0·00877	0·00887	−1·1
50	0·00737	0·00744	−0·9
65	0·00646	0·00644	+0·3
80	0·00554	0·00547	+1·3

TABLE III. (*cont.*).

Thionyl Chloride.

A=0·001874. c=184.

Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
0	0·00814	0·00689	+15·4
15	0·00680	0·00635	+ 6·6
25	0·00603	0·00604	− 0·2
35	0·00577	0·00574	+ 0·5
50	0·00534	0·00537	− 0·5
65	0·00505	0·00504	+ 0·2

Sulphuryl Chloride.

A=0·001184. c=283.

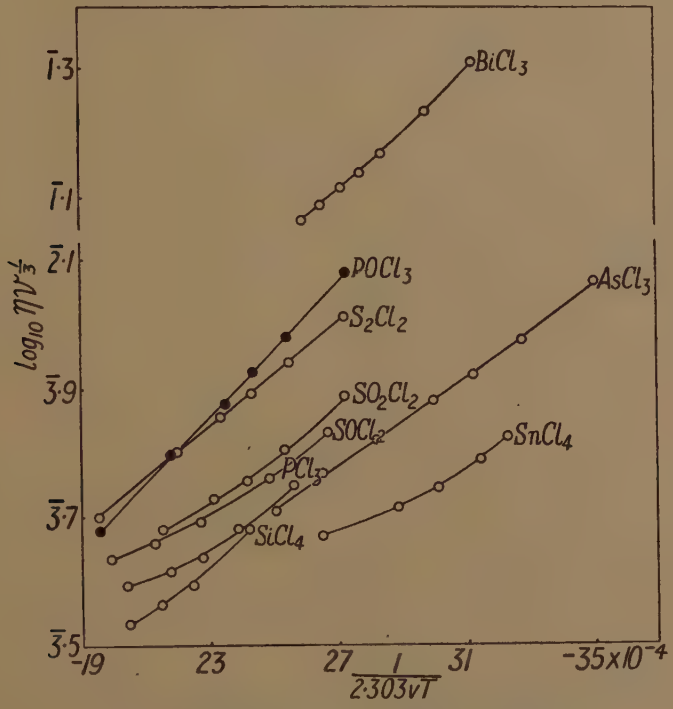
Temp. °C.	$\eta_{obs.}$	$\eta_{calc.}$	Deviation (per cent.).
0	0·00931	0·00832	+10·6
15	0·00762	0·00731	+ 4·1
25	0·00674	0·00676	− 0·3
35	0·00632	0·00627	+ 0·8
50	0·00562	0·00565	− 0·6

The viscosity values for stannic chloride and bismuth trichloride are those given in the 'International Critical Tables.'

The calculations of the constants for all the liquids were based on the curves shown in fig. 2. If the liquid conforms to the Andrade relation, the plot of log $\eta v^{\frac{1}{3}}$ against $1/v\mathrm{T}$ is a straight line. The chlorides of the elements of Group IV. of the Periodic Table appear to show increasing deviation with atomic number and metallic character of the elements. Thus, Andrade (4) has shown that carbon tetrachloride conforms to the relation with a maximum deviation of 1·1 per cent., but silicon tetrachloride and stannic chloride show departures at the lower temperatures. It is clear from the accepted electronic structures of these substances that these deviations cannot be ascribed to simple bonding or to dipolar interaction, since the liquids have zero dipole moment. This feature, however, is not evident in the chlorides of the elements of Group V., for phosphorus trichloride exhibits greater deviation than arsenic trichloride, which, like bismuth trichloride, shows little departure from the Andrade relation. The viscosity of sulphur monochloride has been measured at various temperatures by Trautz (6), but the values do not conform so closely to the above relation as those recorded in the present paper. The departure in the case of thionyl chloride appears to be slightly greater than that of sulphuryl chloride, but it is interesting to note that phosphorus oxychloride conforms to the relation even below the melting-point.

The relevant physical properties of the liquids under investigation are listed in Table IV., where μ is the dipole moment in Debye units, T_c is the critical temperature on the absolute scale, l_F is the latent heat of fusion in cals. per gram, and A and B are the constants of the simple equation (1).

Fig. 2.



The Andrade relation for inorganic liquids.

TABLE IV.

	μ.	T_c.	l_F.	A×10^6.	B.	BR/M.	β/T_B.	β/T_c.
CCl$_4$	0	°556·1	41·1	148	1223	15·8	1·52	0·95
SiCl$_4$	0	503·0	10·9	568	594	6·9	0·78	0·51
SnCl$_4$	0	591·7	8·4	954	691	5·3	0·77	0·51
PCl$_3$	1·2	558·5	—	1357	389	5·6	0·49	0·30
AsCl$_3$	2·15	629·0	—	406	981	10·7	1·06	0·68
BiCl$_3$........	—	—	—	4966	2200	13·9	1·33	—
S$_2$Cl$_2$	—	668·0	—	429	921	13·6	0·97	0·60
POCl$_3$	—	—	20·6	238	1114	14·4	1·28	—
SOCl$_2$	1·58	—	—	135	447	7·5	0·55	—
SO$_2$Cl$_2$	1·86	—	—	647	700	10·3	0·89	—

With these inorganic liquids there is no possibility of direct bonding of the type exemplified in the hydroxyl and hydrogen bonds. In the dipolar liquids, which do not accurately conform to the Andrade formula, the deviation may be ascribed to dipolar association, but it is clear that there is no correspondence between the extent of the departure and the actual value of the dipole moment. In fact, arsenic trichloride with a high dipole moment shows almost complete conformity with the relation, while phosphorus trichloride with a low dipole moment exhibits a departure from the rule. It is customary to attribute the deviations to a "kind of association," but, as Andrade (4) has pointed out, these are to be ascribed, not to the association, but to the change in the degree of association with temperature.

De Guzman (2) and Drucker (3) were the first to show that a relation exists between the constant B of the simple exponential formula and the molecular heat of fusion, and this connexion has been further examined by Iyer (3) on the basis of Raman's theory. According to Iyer the exponential formula may be written

$$\eta = \mathrm{A}e^{\frac{\mathrm{L_F}}{\mathrm{RT}}},$$

where $\mathrm{L_F}$ is the molecular latent heat of fusion and R is the gas constant. A comparison of the values of BR/M (where M is the molecular weight) with the latent heat of fusion reveals a fair correspondence in the case of a number of organic liquids, and the value of B is roughly identified with the latent heat of fusion of the substance. There are, however, several exceptions, and the usual distinction between normal and associated liquids does not hold good.

Since great accuracy is not required for the comparison with the latent heats of fusion, the constants A and B in Table IV. have been estimated from the curves obtained by plotting log η against 1/T. The value of BR/M is materially lower than that of l_F, but the discrepancy is less marked than that found for carbon tetrachloride. In general, the values for the inorganic liquids are appreciably lower than those found by Iyer for organic liquids, but this is to be ascribed to the difference in molecular weight and not to any fundamental structural differences. Approaching the problem from the standpoint of Bernal's configurational theory, Ward (3) has shown that for liquids of the same structural type in which the change from solid to liquid corresponds only to a transition from order to disorder, the value of $\mathrm{BR/L_F}$ is constant. For ionic liquids, such as sodium chloride and potassium chloride, the value is approximately 1·2. Much lower values are found for silicon tetrachloride (0·64) and stannic chloride (0·62), and the exceptional character of carbon tetrachloride (0·38) is again evident.

Applying the exponential equation in the form

$$\log \eta = \alpha + \beta/T,$$

where α and β are constants, Prasad (7) has found that for a large number of organic liquids β/T_B and β/T_C are constants, where T_B and T_C are the boiling-point and critical temperature respectively. For organic liquids, the value of β/T_B is 1·04–1·24 and that of β/T_C 0·64–0·73. Similar values are found for arsenic trichloride, bismuth trichloride, sulphur monochloride, and phosphorus oxychloride, and these are the liquids which are actually found to conform to the Andrade relation over the entire range of temperature examined in the present work.. The other liquids give lower values for both constants. It may be noted that carbon tetrachloride, while conforming to the Andrade relation, gives exceptionally high values which lie well outside the normal range of the Prasad constants.

Recently, Nissan and Clark (8) have pointed out that viscosity may be correlated with the function T/T_B, where T is the temperature at which the measurement is made and T_B is the boiling-point of the liquid. If η_B and η_T are the viscosities at T_B and T respectively, the plot of T/T_B against η_B/η_T leads to a series of fan-shaped curves enclosing regions which correspond to associated liquids, normal liquids, and " metallic " liquids. Employing this method of graphical representation we find that the majority of the inorganic liquids give curves which lie in the central region comprising the normal liquids.

Summary.

The viscosity of some representative non-polar and dipolar inorganic liquids has been measured at a series of temperatures with a view to testing the applicability of the exponential formula. It is found that the majority of the liquids conform fairly closely to the Andrade relation, the largest departures being shown by phosphorus trichloride and stannic chloride. The connexion between the values of the constants of the exponential formula and the other physical properties of the liquids is discussed. In so far as viscosity phenomena are concerned, it is shown that the majority of these substances are normal liquids, which do not suffer change of association with temperature and thus conform to the Andrade relation in the manner of organic liquids.

References.

(1) Arrhenius, *Med. K. Vetenskaps. Nobelinstitut*, iii. no. 20 (1918).

(2) De Guzman, *Annales de la Sociedad Española de Fisica y Quimica*, xi. p. 353 (1913).

(3) Drucker, *Z. phys. Chem.* xcii. p. 287 (1918). Kendall and Monroe, J. Amer. Chem. Soc. xxxix. p. 1799 (1917). Raman, ‘ Nature,’ cxi. pp. 532 & 600 (1923). Frenkel, *Z. Physik*. xxxv. p. 664 (1926) ; ‘ Nature,’ cxxv.

p. 581 (1930) ; *Acta Physicochimica* U.R.S.S. iii. p. 633 (1935). Dunn, Trans. Faraday Soc. xxii. p. 401 (1926). Madge, ‘Nature,’ cxxv. p. 590 (1930) ; J. Phys. Chem. xxxiv. p. 1599 (1930). Sheppard, ‘Nature,’ cxxv. p. 489 (1930) ; J. Rheology, i. pp. 208 & 349 (1930). Andrade, ‘Nature,’ cxxv. pp. 309 & 582 (1930) ; *ibid.* cxxviii. p. 835 (1931) ; Phil. Mag. xvii. pp. 497 & 698 (1934). Iyer, Indian J. Phys. v. p. 371 (1930). Prasad, Phil. Mag. xvi. p. 263 (1933) ; J. Indian Chem. Soc. x. pp. 135 & 143 (1933) ; *ibid.* xi. p. 913 (1934). Silverman, Trans. Faraday Soc. xxix. p. 1285 (1933). Eyring, J. Chem. Phys. iv. p. 283 (1936) ; *ibid.* v. p. 726 (1937). Macleod, Trans. Faraday Soc. xxxii. p. 872 (1936). Bernal, *ibid.* xxxiii. p. 27 (1937). Ward, *ibid.* xxxiii. p. 88 (1937). Van Wijk and Seeder, *Physica*, iv. p. 1073 (1937). Simha, J. Chem. Phys. vii. p. 202 (1939).

(4) Andrade, Phil. Mag. xvii. p. 698 (1934).

(5) Barr, ‘Monograph of Viscometry’ (Oxford, 1931). Hatschek, ‘The Viscosity of Liquids’ (Bell, 1928). Bingham, ‘Fluidity and Plasticity’ (McGraw-Hill, 1922).

(6) Trautz, *Z. Elektrochem.* xxxv. p. 130 (1929).

(7) Prasad, J. Indian Chem. Soc. x. p. 135 (1933).

(8) Nissan and Clark, ‘Nature,’ cxliii. p. 722 (1939).

XXXVIII. *The Rôle of Ionization by Positive Ions in Spark Breakdown.*

By R. N. VARNEY, Ph.D., Assistant Professor of Physics, Washington University, St. Louis, Mo., L. B. LOEB, Ph.D., Professor of Physics, University of California, Berkeley, Calif., and W. R. HASELTINE, Ph.D., Research Associate in Physics, University of California, Berkeley, Calif.*

[Received October 20, 1939.]

IN 1900, Prof. J. S. Townsend (1) first published his famous theory of ionization by collisions of ions which led to a mechanism of spark breakdown. According to the theory, the necessary ionization for such breakdown in a gap was produced by the collisions of electrons (at the time thought to be negative gas ions) and of positive ions. In the gap before breakdown occurred, a certain initial photo-electric current i_0 from the cathode grew by these ionizing collisions into a current i at the anode given by the equation.

$$i=i_0\frac{(\alpha-\beta)e^{(\alpha-\beta)d}}{\alpha-\beta e^{(\alpha-\beta)d}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Here α is the coefficient of ionization by electrons, β is the coefficient of positive ion ionization, and d is the anode-cathode separation. The coefficients α and β divided by the pressure p were shown to be functions of X/p, the ratio of electric field strength to pressure only.

Experiments showed that the current i satisfied such an equation over a certain range provided α and β were properly chosen. It was used extensively as a method for determining the values of α and β. In this equation it is seen that when $\alpha=\beta e^{(\alpha-\beta)d}$, i becomes indefinitely large irrespective of i_0. This Townsend set as the criterion for a spark. And, in fact, the condition fulfils Paschen's law and leads to a prediction of the value of V_s, the minimum sparking potential *at low values* of $p.d$ above the minimum (up to $p.d\sim 100$) but far below breakdown at atmospheric pressure, in fair agreement with experiment.

The early agreements at high X/p and low values of $p.d$ led Townsend to believe that this constituted a proof that the assumed mechanism was truly the spark breakdown mechanism at all pressures and therefore that ionization by positive ions occurred. This conclusion was strongly supported by the appearance of positive corona discharges where the cathode could not play a part.

---

* Communicated by the Authors.

It was first pointed out by Holst and Oosterhuis (2) that for low pressures V_s was dependent on cathode material which is not consistent with an ionization by positive ions, and that, furthermore, in the fields existing, positive ions could not get the energy to ionize by impact. The dependence on cathode material has since been confirmed by Neuman (3) and by Ehrenkranz (4). Subsequent studies on ionization by positive ions have borne out this conclusion, as will later be shown. Further experiments and theoretical analyses have shown that the coefficient β could be interpreted as any one of a number of other processes (5). Thus, if it is assumed that the positive ions liberate electrons from the cathode by collision, that photo-ionization occurs in the gas, that further photo-electrons are liberated from the cathode by the radiation from the discharge, etc., and a corresponding equation derived, it is found that an equation of form differing slightly and experimentally indistinguishable from that of equation (1) is obtained containing a coefficient corresponding to β. In addition, more recent work has shown that positive corona can be accounted for by photo-ionization in the gas (6).

Despite these developments, Prof. Townsend (7) has recently published a paper in which he purports to prove the possibility of ionization by positive ion impact. Inasmuch as some of the data and conclusions used by Townsend are not in agreement with modern physical interpretation and fact, and because of the great importance to physicists and engineers of a correct understanding of the spark breakdown problem at atmospheric pressure, it seems desirable to correct the erroneous conclusions.

Townsend first uses the criterion for a spark that the denominator of equation (1) must become zero, that is,

$$\beta=\alpha e^{-(\alpha-\beta)d}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

or in terms of secondary liberation at the cathode by positive ions or photons that

$$\gamma=e^{-\alpha d}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where γ is the coefficient of liberation of electrons from the cathode by positive ions or by photon bombardment. Now an analysis by Loeb (8) shows that this equation as applied as a criterion for sparking is mathematically and physically meaningless. Based on the criterion of a self-sustaining discharge, initially due probably to Holst and Oosterhuis (2) or Seeliger, the same condition can be set up as the minimum value of d for a self-sustaining discharge. This, by establishing a threshold of instability, sets a minimum value for the field strength necessary for a change in the current characteristics in the gap, which may initiate a spark. It is essential to recognize to-day, however, that despite the existence of the minimum threshold, an exact and precise breakdown potential as inferred from Townsend does not even exist, and that for gaps greater than 3 mm. long and at more than 100 mm. pressure in molecular gases, breakdown is a chance phenomenon whose frequency of occurrence may be predicted by probability calculations as a function of voltage but cannot be laid down in any absolute sense. At lower pressures and gap lengths, the modification of Townsend's original theory redefining the meaning of β does apply on the basis of the minimum above and fairly well approximates the results. The mechanism that actually occurs above a 3 mm. gap length and 100 mm. pressure is a quite different one, depending on streamer formation and photo-ionization in the gas, as shown in the very recent work of Meek (9) shortly to be published. That anything like a fixed potential appears to exist is the result of the very rapid change in probability near the critical region.

In applying the breakdown criterion to calculate β at *atmospheric pressure*, Townsend assumes that at atmospheric pressure, with a 1 cm. long gap, the breakdown potential will be 33,000 volts. Using this value, he finds from accepted experimental measurements that α is 19·8 and therefore that β is $19{\cdot}8e^{-19{\cdot}8}$, or approximately $5 . 10^{-8}$. Actually, the data on the breakdown potential are highly discordant, for the reason mentioned above, and it is impossible with existing data and with the varying materials and techniques used to pick with certainty the most appropriate values. In general, with reasonable ultraviolet illumination of the cathode the minimum value observed should be chosen. In working up their theory, Meek and Loeb (9) agreed that the value of 31,600 volts given by Schuman and by Spath (10) is perhaps the best value to use. Thus the value of 33,000 volts is probably excessive. While a seemingly trivial point, the fact that this sets the value of an exponent which profoundly affects the numerical values requires that some care be used in the adoption of these values. The value of 31,600 used by Meek makes α about 17. The value of β is thus altered by a factor $e^{2{\cdot}8}$ or 16·5, a whole order of magnitude. β should probably be $7{\cdot}1 . 10^{-7}$ instead of $5 . 10^{-8}$.

Townsend next undertakes the calculation of the number of positive ions in the gap which succeed in acquiring an energy sufficient for ionization. In the calculation three assumptions are made: (1) It is assumed that air of ionization potential 15 volts may be ionized by positive ions of energy 30 volts. (2) It is assumed that the mean free path of an ion can be correctly calculated by an extrapolation of the viscosity theory to extremely high temperatures. (3) It is assumed that the fields are such that the free paths lie entirely in the field direction. The two constants thus obtained, V_0 and λ, are then applied in the gas kinetic free path equation

$$n=\frac{d}{\lambda}e^{-\frac{V_0 d}{V\lambda}}, \quad \dots\dots\dots \quad (4)$$

where n is the number of ions which travel such a distance as to acquire energy at least V_0, d is the gap length, V is the breakdown potential, and λ is the mean free path of a positive ion. The critical quantities all occur in the exponent. It must be emphasized again that the greatest precautions are necessary where exponentials are involved, for a seemingly trifling error in one number may be multiplied thousands of times. Thus, for instance, Townsend's calculation of n gives

$$n=14{,}300e^{-13}.$$

If λ is in error by a mere factor of 2 in this equation, n is altered by a factor of $2.e^{-13}=5.10^{-6}$.

It appears certain to the authors that not only are both V_0 and λ as used in error by a very considerable factor and that both errors militate against the very occurrence of ionization by positive ions, but that, in principle, the calculation of β by this equation is erroneous.

(1) *No direct experimental evidence exists* in which ionization of a gas by 30 volts positive ions occurs. Furthermore, such direct evidence has been intensively sought. Eleven years ago, Loeb (11) pointed out the existence of two distinct mechanisms of ionization by collisions of particles heavier than electrons, and these have been carefully studied since then. On the one hand, ionization by α particles, high speed protons and deuterons, etc., such as occurs in Wilson cloud chambers, has been shown by Döpel (12) to be essentially a process depending on the high velocity of the ionizing particles. Theoretical treatment of this problem is considered to be satisfactory. But Döpel was unable to observe even excitation, let alone ionization, at energies below 2000 volts in such collisions. In other words, this high energy process has nothing whatever to do with spark breakdown phenomena.

On the other hand, a certain peculiar resonance ionization phenomenon seems to occur in the atomic gases like Ne, A, Kr, Xe, Hg (and perhaps in H_2) when the atoms are struck by alkali ions or by their own fast neutral atoms. Ionization by alkali ions appears much like the ionization of these gases by electrons. The ionization efficiency is quite high, exceeding one per cent. in several cases, and is characterized by a sharp onset. Varney (13) has observed ionization to begin when alkali ions traverse these gases at the energies listed in Table I.

It is to be noted that all these onsets occur at energies not twice the ionization potential but at *five* or *six* times the ionization potential. No ionization potential was observed below 500 volts by any alkali ions in H_2, He, N_2, CO_2, or CH_4, nor was ionization observed in any of the spaces marked with a dash in Table I.

Related to this ionization by alkali ion collisions is the ionization produced in these same gases by their own fast neutral atoms. When ions, for example, of argon are allowed to pass through argon gas, many of the ions pick up an electron from a neutral atom by a resonance exchange mechanism and proceed with unchanged velocity as neutral atoms. When these neutral atoms collide with stationary atoms, they ionize much as the alkali ions do. Varney (14) has observed the same sort of sharp onset to occur in these cases listed in Table II.

Here, too, the efficiency is high, even higher than for alkali ions. Attention is again called to the fact that ionization does not begin at twice the ionization potential, but at more than three times that value. No such ionization was observed in N_2 up to 200 volts. Döpel (12) has observed excitation in H_2 by neutral H atoms at energies of 500 volts.

TABLE I.

Onset Potentials in Volts.

(True ionization potentials are added for reference.)

	Ne.	A.	Kr.	Xe.	Hg.
Na	130	—	—	—	80 ca.
K	—	82	69	114	—
Rb	—	135	97·5	150	—
Cs	—	338	200	77	—
Electrons	21·5	15·5	13	11·5	10·5

TABLE II.

	Onset potentials in volts.	Ionization potentials.
Ne–Ne	74	21·5
A–A	48	15·5
Kr–Kr	40	13
Xe–Xe	35	11·5

If ionization by positive ions occurs in spark breakdown, it must certainly occur by this process of neutralization first and then by ionizing collision. If direct positive ion collisions could produce ionization, more energy would necessarily be required, for the detached electron would have to escape from two positive charges instead of from one as in the neutral atom case. The following questions then arise concerning this mechanism :—(*a*) Does it occur in other cases than those listed? (*b*) If so, how efficiently can it occur and still escape detection as reported above ? (*c*) At what energies may onset be anticipated ?

(*a*) The entire mechanism involved here appears to be one peculiar to certain gases and certain ions. The atomic nature of the gases or their completed external electron shell may influence the mechanism. *There are no satisfactory observations of such ionization occurring in any other cases.* Furthermore, the few observations that exist only claim to show ionization at much higher energies, greater than 100 to 200 volts.

(*b*) Beeck (15) has made quantitative measurements on the ionization collision cross-section for ionization by alkali ions. He states, for example, that if one K^+ ion in roughly 10,000 at 500 volts energy ionized helium he could have detected it. Various comparisons of the work on the balanced space charge ion detectors done by Varney (16) with the absolute measurements of Beeck and of Rostagni (17) indicate that under the same circumstances, ionization to the extent of one in 50,000 to 100,000 could have been detected. A further point is to be noticed in this connexion. The ionization function, *i. e.*, the curve of ionization efficiency against energy of ionizing particle, must necessarily have some positive slope. If, then, no ionization is observed at 500 volts to one part in 10^{+4} or 10^{+5}, the chance that ionization occurs at energies of 50 volts is doubtless less than one in 10^5 or 10^6, and probably very much less. It seems most unlikely then that ionization in air even as feeble as that demanded by Townsend's numerical interpretation of his theory could escape direct detection.

(*c*) Townsend has assumed that positive ions with energy twice the ionization potential of the gas could produce ionization. The results cited above contain two contradictions to this assumption. First, in even the most efficient cases, ionization was never observed to set in at less than three times the ionization potential. Secondly, it has been generally observed that there is a qualitative relationship between efficiency of ionization and height of the onset potential ; the less efficient the ionization the higher the onset. Hence the known feebleness of any ionization not observed by Beeck or Varney indicates a much higher onset. While this point may not be applicable, the first must be reiterated, namely, that *no ionization has ever been observed at less than three times the ionization potential.* If these data, based on the most careful modern work, are accepted, then it is necessary to suppose that *no less than* 45 volts and probably 50 volts would be required for ionization by positive ion collisions in air. Since the voltage enters into the exponent, an error of 20 volts would change Townsend's results by a factor of nearly e^{-9} or about 10^{-4}, that is, the chance that positive ions would get enough energy to ionize would fall off by this factor of 10^4.

(2) The second of Townsend's assumptions, namely, that the ion free path may be calculated by an extrapolation of viscosity theory to high temperatures, results in even more serious experimental disagreements. One might be inclined to question the validity of the extrapolation to higher temperatures on the grounds that the *decrease in molecular size does not continue indefinitely*. But there are many experimental results for ions in gases which show that the normal ion mean free path, far from being longer than the molecular mean free path, is, on the contrary, from one-third to one-fifth as long at lower energies and *reaches a limit barely equal to the molecular path at the higher energies* of 20 to 30 volts. This conclusion was first indicated in theory and experiment by the results of mobility and diffusion rates of ions. Hershey (18) has measured the mobility of K^+ ions in H_2, N_2, He, and A at low pressures at values of X/p from virtually zero to 200. He finds that the value of the mobility, and hence of the ion mean free path, at first increases very slightly and then drops as X/p is increased. Theoretical comparison agrees excellently provided the forces of attraction arising from polarization of the molecules by the charged ions and from van der Waals's theory are included. Omission of these forces is equivalent to an elastic solid theory and approaches closely to the case of the ordinary molecular collisions. Such an omission gives a mobility approximately three times the correct value at low X/p, indicating therefore that the ion mean free path is less than the molecular mean free path to this extent.

At higher X/p the forces become less effective because of the increased velocities. Hershey's experiments agree excellently here too with the theory he has developed, based on a solution of Boltzmann's equation, leaving no doubt whatsoever as to the correctness of the decreased mobility value observed at high fields and lower pressures.

Further evidence is due to Ramsauer and Beeck (19), who measured the effective radii directly of He, Ne, A, N_2, O_2, H_2 when the ions Li^+, Na^+, K^+, Rb^+, and Cs^+ were passed through them with energies from about 1 to 30 volts. The observed radii in every case were larger for the slower ions, as indicated by the mobilities, than the kinetic theory radii,and approach near to the kinetic theory values at voltages greater than 20. While in three cases (Li^+ in He and in A, Cs^+ in He) the radii eventually became very slightly smaller than those predicted by kinetic theory, they all were very near to the kinetic theory values at the higher energies, and most were actually larger by a factor not exceeding 1·5. The mean free path of the fast ions is thus very nearly the same as that for molecules or, if anything, *smaller*, certainly not 7·3 times larger, as Townsend concludes. There is no evidence of an effect for ions analogous to the Ramsauer (20) effect for electrons giving anomalously high free paths, nor would theory lead one to expect it. The only theoretical indication of any material difference comes in wave mechanics for the interaction of He^{++} and He where the electrons are only two in number and the ion is bare of electrons. This does not apply to more complex systems. The only significant difference between alkali ions and actual gas ions is that the latter are subject to the Kallman and Rosen (21) charge exchange, which even further shortens the mean free path.

Barring other errors in the mean free path * and disregarding the error in V_0, the minimum ionizing energy, and in V, the breakdown potential, the calculation of the number of times an ion acquires more than 30 volts energy in crossing the gap given by Townsend should be

$$n=\frac{1}{9.10^{-6}}e^{-\frac{30\cdot 1}{33000.9.10^{-6}}}=2{\cdot}8\,.\,10^{-38}$$

This differs from Townsend's value of n by assuming that $\lambda_{(\mathrm{ion})}=\lambda_{0(\mathrm{gas})}$ instead of $\lambda_{(\mathrm{ion})}=7{\cdot}3\,.\,\lambda_{0(\mathrm{gas})}$, and the result differs by a factor of 10^{-36}. According to this corrected result, an ion would need to produce $3{\cdot}3\,.\,10^{31}$ ions by collision each time it had more than 30 volts energy, an obviously impossible and ridiculous result.

There are inherent in the calculations of Townsend two more considerations that have been neglected which might reduce the absurdity of the conclusions above, but not materially. In the first place the persistence of velocities leads one to expect that the energy of the ions in the field may be greater than $Xe\lambda$ that gained in one free path, which might favour his calculation. If an attempt is made to correct his procedure for this fact, one finds that one runs into a logical inconsistency which will become clearer later. Again, Townsend set the collision frequency as $1/\lambda$, which on a strict interpretation of his theory is consistent but which is again entirely contrary to fact. These two corrections to his theory, increasing the energy of the ions and increasing the collision frequency which would yield larger values of β, however, indicate that the whole procedure used by Townsend in calculating β is physically incorrect.

The reason that a calculation such as Townsend's is physically incorrect lies in his assumption (3) that the ion free paths in fields of this order of magnitude will all be directed along the field axis. An exactly similar mistake was made by Townsend in calculating α for ionization by electron impact, only in the latter case because of elasticity of electron impacts the result was even more seriously in error. The ions do not lose all their energy gained in a free path at each collision, in fact they rebound with

---

* As indicated in (3), Townsend has assumed that the free paths are all exactly in the field direction. This is not the case. A correction factor for this, even if only 0·7, would further enormously decrease the probability because of the size of the exponent.

a considerable fraction of their energy in all directions. Unless the fields are of such a magnitude that the randomly rebounding ions are oriented in the field direction, one cannot therefore use the mean free path expression alone for determining the probability of the acquisition of ionizing energy, for that expression must then be multiplied by the probability that the ion path as a result of its random thermal velocity will be so directed as to have its component in the field direction equal to λ. Such a calculation might be made if the average energy of the ions were known but would be difficult and tedious. It is obvious that the only correct and sensible procedure is to use calculations analogous to the form of the calculations successfully applied to the value of Townsend's α in a simple way by Emmeleus, Lunt, and Meek (22) or von Engel and Steenbeck (23), or the more esoteric calculations made by the method of Druyvesteyn (24). It is possible to indicate the order of magnitude of the probability of ionization by K^+ ions in N_2, for example, in an elementary fashion with existing data on hand using the concepts indicated above based on the energy distribution. For this purpose, we will be forced to use a Maxwellian form of energy distribution for positive ions which the frequent and heavy energy exchanges with gas molecules make relatively plausible and which has been recently employed by Hershey for ion mobilities at higher X/p with some success.

In order to calculate the probability of ionization of K^+ in N_2, we will then set up a Maxwellian velocity distribution in which, however, we will use not the room temperature but a raised temperature to correspond to the mean energy which the ions still possess due to the field. A further correction to the distribution law will be made to set the distribution relative to the drift velocity instead of to zero.

To determine the mean temperature of the ions, we will first find a mean velocity from Langevin's mobility equation

$$k=0{\cdot}815\frac{e}{m}\frac{\lambda}{C}\sqrt{\frac{m+M}{M}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where k is the mobility, e the electronic charge, m the mass of the ion, M the mass of the molecule, and λ the mean free path. Substituting $k=3$ cm./sec./volt/cm., $e=4{\cdot}8.10^{-10}$ esu, $M=28.1{\cdot}66.10^{-24}$ gm., $m=39.1{\cdot}66.10^{-24}$ gm., and $\lambda=10^{-5}$ cm., a value of C is calculated to be $1{\cdot}04.10^5$ cm./sec. Hence the kinetic energy, $1/2\ mC^2$, becomes $0{\cdot}350.10^{-12}$ erg, and kT is two-thirds of this or $0{\cdot}233.10^{-12}$ erg, which in turn is 0·146 volt. The corresponding temperature is 1700° K. Calculations made on the basis of Hershey's work in which forces between the ions and molecules are taken into consideration give a result nearer 1400° K. Both of these are nearly in agreement with experimental measurements of plasma temperatures in discharges.

To account for the distribution relative to the drift velocity, we will use a distribution law of the form

$$f=\frac{1}{(2\pi m_1 kT_1)^{3/2}}e^{-\frac{(p-m_1v)^2}{2m_1kT_1}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where kT_1 is to be used as 0·51 volt, p represents the vector momentum, and m_1v is the drift momentum.

The value of β may then finally be calculated from the fact that the number of ionizing collisions per cm. path per ion is equal to the number of collisions per cm. times the chance of acquiring the minimum necessary energy (which we will still take as 30 volts) times the chance that ionization will occur at a single collision at the necessary energy (represented by P in our equation).

$$\beta=\frac{P}{\lambda v}\frac{2}{\pi^{1/2}a^3}\int_V^\infty\int_0^\pi e^{-\frac{(C^2-2Cv\cos\theta+v^2)}{a^2}}C^3\sin\theta\,d\theta\,dC. \quad . \quad . \quad . \quad . \quad (7)$$

Evaluation of the integrals has been somewhat simplified by omission of a very small term ; since the term had a negative sign we have again erred on the side favouring Townsend's interpretation. The result may be written

$$\beta=\frac{P}{2\pi^{1/2}\lambda}\frac{a}{v^2}\left[(V+v)+\left(\frac{a^2}{2}+v^2\right)\frac{1}{(V-v)}\right]e^{-\left(\frac{V-v}{a}\right)^2} \quad . \quad . \quad . \quad (8)$$

The second term in the bracket is only of the order of one one-hundredth of the first term and may be neglected if desired. The effect of considering the thermal motion of the gas molecules is to replace a^2 by $a^2+a_0{}^2$ or T by $T+T_0$, where T_0 is the temperature of the gas. Thus the a^2 to be used corresponds to 1700°+300°=2000° or 0·171 volt. For good measure we shall use 3000° or 0·257 volt. Then for $k=3$ cm./sec./volt/cm., $X=33{,}000$ volts/cm., $v=kX=\sqrt{0{\cdot}191}$ volt, $a^2=0{\cdot}257$ volt, $V=\sqrt{30}$ volts, $\beta=5\,.\,10^{-8}$, $\lambda=10^{-5}$ cm., we find that P must have the value $9{\cdot}2\,.\,10^{29}$. Here again, in other words, the conclusion is that each ion with 30 or more volts energy must produce 10^{30} ions at each collision.

Numerical experimentation with this equation shows that the mean ion energy would have to be in excess of 6 volts if the value of P were taken as 10^{-6} and V were set at 50 volts in accordance with the observations previously described. As a matter of fact, it has already been pointed out that the energy of one-quarter volt we have used is itself too large.

In conclusion, it can be seen that (1) even on the basis of Townsend's erroneously based theory of ionization by positive ions, when the modern values of the constants used are inserted, his equations lead to probabilities of ionization which are absurd. (2) When one considers the proper mode of procedure following classical kinetic theory lines laid down by Boltzmann and others, one likewise finds that the probability of ionization by positive ions under existing sparking conditions is absurdly small. One is therefore forced to the conclusion that, in conformity with the direct measurements, ionization by positive ions in a gas under ordinary sparking conditions and even at low pressure and high X/p is virtually ruled out. There is in principle no great difficulty in calculating the value of β for positive ions in a gas following the procedure of Emmeleus, Lunt, and Meek (22), if one assumes the Maxwellian law, which is obviously not far from the truth, since there are no Ramsauer free paths for ions as for electrons. The mathematical difficulty in solving the Boltzmann equations for the distribution of energy among positive ions in a field in order to obtain the exact form of the law presents at present formidable mathematical difficulties. Thus the exact solution of the problem must lie in the future. Even with the approximate solutions, however, it is seen that the discrepancy between values of β measured and those calculated is such as to make it unlikely that ionization by positive ions can be demonstrated theoretically.

References.

(1) J. S. Townsend, 'Electricity in Gases.' Oxford, The Clarendon Press, 1915.
(2) Holst and Oosterhuis, Phil. Mag. xlvi. p. 1117 (1923).
(3) L. J. Neuman, Nat. Acad. Sci. Proc. xv. p. 259 (1929).
(4) F. Ehrenkranz, Phys. Rev. lv. p. 219 (1939).
(5) See, for example, L. B. Loeb, Rev. Mod. Phys. viii. p. 267 (1936).
(6) Loeb and Kip, Jour. App. Phys. x. p. 142 (1939) ; G. W. Trichel, Phys. Rev. lv. p. 382 (1939) ; A. F. Kip, Phys. Rev. lv. p. 549 (1939).
(7) J. S. Townsend, Phil. Mag. xxviii. p. 111 (1939).
(8) L. B. Loeb, 'Fundamental Processes of Electrical Discharge in Gases.' John Wiley & Sons, New York, 1939, p. 423 ff.
(9) Submitted to the Phys. Rev., August 1939.
(10) W. D. Schuman, 'Elektrische Durchbruchfeldstärke von Gasen.' Berlin, J. Springer, 1923 ; W. Spath, *Arch. f. Elektrot.* xii. p. 331 (1923).
(11) L. B. Loeb, 'Science,' lxvi. p. 627 (1927).
(12) R. Döpel, *Ann. d. Phys.* xvi. p. 1 (1933).
(13) R. N. Varney, Phys. Rev. xlvii. p. 483 (1935) ; Varney and Cole, Phys. Rev. l. p. 261 (1936) ; Varney, Cole, and Gardner, Phys. Rev. lii. p. 526 (1937).
(14) R. N. Varney, Phys. Rev. l. p. 159 (1936) ; see also N. Wayland, Phys. Rev. lii. p. 31 (1937).
(15) O. Beeck, *Ann. d. Phys.* (5) vi. p. 1001 (1930) ; *Phys. Ztschr.* xxxv. p. 36 (1934).
(16) R. N. Varney, Phys. Rev. liii. p. 732 (1938).
(17) A. Rostagni, *Nuovo Cimento*, xi. p. 34 (1934) ; Phys. Rev. liii. p. 729 (1938).
(18) A. V. Hershey, Ph.D. thesis, University of California, Sept. 1938, Phys. Rev. lvi. p. 908 (1939).

(19) Ramsauer and Beeck, *Ann. d. Phys.* (4) xlvii. p. 1 (1928).
(20) See, for example, the summary article by R. B. Brode, Rev. Mod. Phys. v. p. 257 (1933).
(21) Kallmann and Rosen, *Ztschr. f. Phys.* lviii. p. 52 (1929) ; lxi. p. 61 (1930) ; lxiv. p. 806 (1930) ; lxxvi. p. 213 (1932).
(22) Emmeleus, Lunt, and Meek, Proc. Roy. Soc. A, clvi. p. 394 (1936).
(23) von Engel and Steenbeck, *Elektrische Gasentadungen*, i. pp. 89, 97, 184, 192. Julius Spring, Berlin, 1932.
(24) M. J. Druyvesteyn, *Physica*, iii. p. 65 (1936).

XXXIX. *Experiments on the Action of Ionizing Radiations on the Electrophoretic Mobility of Colloid Particles.*

By J. A. Crowther, M.A., Sc.D., Professor of Physics in the
University of Reading, H. Liebmann, Ph.D.,
and R. Jones, B.Sc.*

[Received December 27, 1939.]

In previous communications from this laboratory (1) it has been demonstrated that the electrophoretic mobility of colloidal particles undergoes rhythmic increases and decreases when exposed to a steadily increasing dose of ionizing radiation. A typical graph of the relation between the mobility and the dose is given in the accompanying figure, which shows the relation between the electrophoretic mobility and dosage for a gold sol of particle radius $2{\cdot}47 \times 10^{-6}$ cm. when exposed to gamma radiation from radium.

It will be seen that the graph shows a series of sharply defined maxima, so sharp, in fact, that the experimental points are best represented by regarding them as definite discontinuities in the curve. The first minimum is equally well defined; subsequent minima, however, appear to be broader. This is possibly more apparent than real. It has already been shown (2) that a gold sol which has received a dose of radiation corresponding to a minimum on the dose/mobility curve is no longer stable, though its stability can be restored by further exposure to radiation. Unless this further irradiation is given the sol will coagulate completely if allowed to stand for a few hours. It is thus possible that during the time which elapses between the irradiation and the mobility measurements there is a certain precipitation of particles taking place which would tend to make the minimum less well defined.

The ascending and descending branches of the curve can be represented, within the limit of the experimental error, as straight lines over the greater part of their length. The experiments, however, are not accurate enough to preclude the possibility of some slight curvature.

Experiments described in a previous paper (1c) indicated that the dose required to produce a certain change in the mobility of a sol varied very rapidly with the radius of the particles (approximately as the inverse square of the particle radius) and indicated the importance in future work of using only homodisperse sols of known particle size. The investigations of the present paper have been made on a series of sols prepared from a seed sol by a technique described by Westgren (3). This method yields a series of very homogeneous sols. The particle number of the sols was found by counting in the ultramicroscope, and the gold content was determined by titration with KCN (4). The gold particles are presumably small octahedra. The radius, a, quoted in the tables is the radius of the sphere which would have the same mass as the actual particle.

---

* Communicated by Professor J. A. Crowther, M.A., Sc.D.

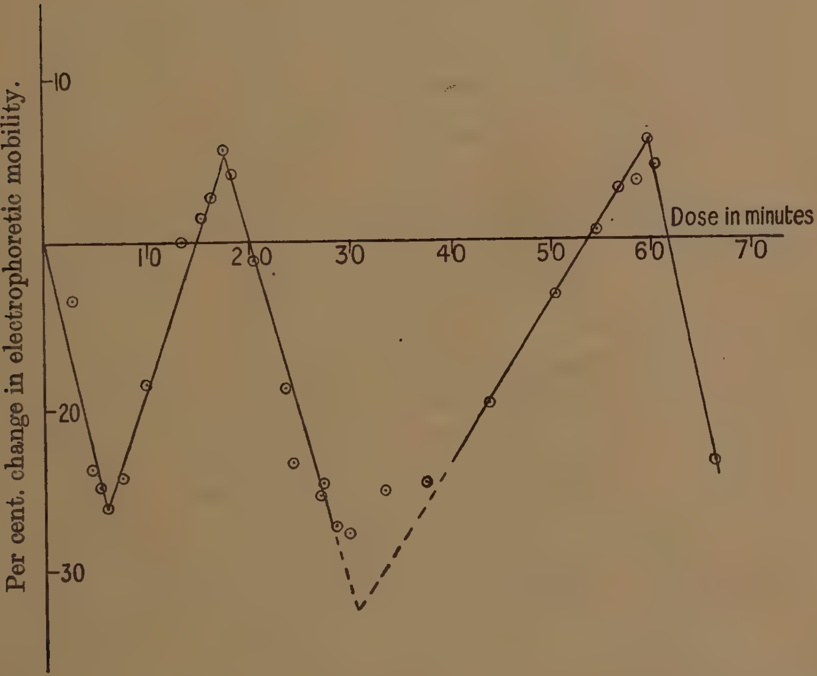
Exposure to gamma radiation was made at a fixed distance from a 50 mgm. source of radium contained in a glass tube and screened with 1 mm. of silver. In order to fix the distance with suitable precision the dish containing the sol and the radium tube were enclosed in a lead block, which also served as protection against undue exposure of the workers in the laboratory. The dose was assumed to be proportional to the time of exposure, and the doses are recorded in minutes. From the known value of the Sievert unit and the geometry of the exposure chamber it was calculated that the direct radiation received by the sol was 2·1 röntgen per minute. Owing to the proximity of the lead walls, however, there must have been a very considerable amount of secondary gamma and beta radiation falling on the sol, which would materially increase the dosage rate. Experiments to determine the actual dosage rate were not completed owing to the necessity of removing the radium to safe storage at the outbreak of hostilities.

In a further series of experiments the sols were irradiated by the fluorescent X-radiation emitted by radiators of molybdenum and selenium, which were illuminated by the direct beam of a X-ray tube. The emission consists almost entirely of the K lines of the radiator, the K_α lines predominating, and provides a source of radiation which is approximately homogeneous. The diffuse radiation from the radiator was limited by two circular stops, so that it just filled the dish in which the sol was to be exposed. A photographic plate in the position of the surface of the sol showed that over this area the beam was of uniform intensity. The radiation reaching the sol after passing through the stops was of rather low intensity. In order to avoid unduly long exposures, which might lead to some precipitation of the sol when passing through its minimum value of mobility, a sol of large particle size ($a=6{\cdot}4\times10^{-6}$ cm.) was prepared for these experiments. The radiation reaching the sol was measured by a flat ionization chamber connected to a counter (5), which in turn was calibrated by a second ionization chamber connected to a condenser and an electrometer. This second chamber could be placed in exactly the position occupied by the sol during exposure.

The results of these experiments are collected in Table I.

The different turning-points in the curve connecting the electrophoretic mobility with the dose are numbered in order : the first minimum being 1, the first maximum 2, and so on. D_n is then the dose of radiation required to bring the sol to the corresponding position. Tabulating the results in this way immediately revealed a marked regularity in the measurements. For a sol or a given particle size each critical dose D_n was approximately twice the one immediately preceding it in the table. The dose required to reach the nth turning-point could thus be derived by a relation of the form

$$D_n \propto k^n,$$



where k is a constant somewhat smaller than 2. From the known values of n and $\log D_n$ the exact value for k could be calculated. For the five groups of measurements given in Table I. the following figures were obtained :—

1·82, 1·71, 1·83, 1·85, 1·89,

with a mean of 1·82. The last column of Table I. indicates the accuracy with which the formula fits the experimental data. Only in the case of the first minimum of the mobility/dose curves does the value of D_n/k^n

differ by more than a few per cent. from the mean value for the given sol. This discrepancy cannot wholly be assigned to experimental error, and it seems clear that the initial part of the curve is somewhat anomalous and cannot be represented by the equation.

Experiments in which colloidal graphite was irradiated with molybdenum K radiation showed that the suggested law applied equally well to this sol (Table II.). The calculated value for k, 1·77, agrees well with those found for gold sols. The constant in the suggested equation, therefore, does not seem to depend appreciably on the nature of the dispersed phase and the nature of the radiation employed. From the form of the equation k must be a non-dimensional constant.

TABLE I.

Radiation.	$a \times 10^6$.	n.	D_n (minutes).	$D_n/1{\cdot}82^n$.
γ	5·78	1	1·3	0·72
		2	2·4	0·72
		3	4·5	0·75
		4	8·0	0·73
	3·47	1	4·0	2·20
		2	6·4	1·92
		3	11·0	1·82
		4	20·0	1·83
	2·47	1	5·6	3·18
		2	18·0	5·41
		3	32·0	5·30
		4	60·0	5·47
			(röntgen)	
$\mathrm{Mo_K}$	6·40	1	2·0	1·10
		2	4·7	1·42
		3	8·2	1·36
		4	14·2	1·30
$\mathrm{Se_K}$.........	6·40	1	1·6	0·88
		2	3·6	1·08
		3	6·4	1·06
		4	11·1	1·01

Experiments on gold sols. a=particle radius, D_n=dose corresponding to nth turning-point of the mobility/dose curve.

In a previous series of observations (1c) it was shown that the dose required to bring the sol to a particular point on the dose-mobility curve varied roughly as the inverse square of the particle radius. The mean value of D_n/k^n for the three gold sols irradiated with gamma radiation is given in the second column of Table III. (In calculating this mean the values for D_1 which appear to be anomalous have been omitted.) The figures for a^2D_n/k^n appear in the last column of the table.

The accuracy of the values of the last column depends on the precision with which the size of the particles can be determined. This cannot be measured with the same degree of accuracy as that obtainable in the measurements of doses and electrophoretic mobilities. The error in the particle size was estimated to be not greater than 5 per cent. This would account for a variation of 10 per cent. in the last column of Table III. The deviations from the mean value, however, are distinctly greater.

TABLE II.

Radiation.	n.	D_n(röntgen).	$D_n/1{\cdot}82^n$.
Mo_K	1	8·4	4·61
	2	15·0	4·51
	3	26·5	4·44

Colloidal graphite.

TABLE III.

$a\ 10^6$.	D_n/k_n.	$a^2D_n/k^n\ 10^{12}$.
5·78	0·73	25
3·47	1·86	23
2·47	5·39	33

Gold sols exposed to gamma radiation.

It must therefore be concluded that either the product is influenced by factors not taken into account (*e. g.*, electrolyte content), or that the dose varies rather more rapidly than $1/a^2$. It may be significant that in the present as well as in the previous series of experiments the values of D_1 varied more rapidly, and those of D_2 and higher critical doses varied less rapidly, than the inverse square of the radius.

The investigations with homogeneous X-radiation were undertaken to determine how, if at all, the phenomenon was affected by the wave-length of the radiation. It was thought that a wave-length effect would most likely occur in the neighbourhood of an absorption band of the material dispersed, and the two radiations employed were chosen to bracket as closely as possible the L absorption band of gold. The relevant data are as follows :—

Gold absorption edges ..	L_1 0·862,	L_2 0·900,	L_3 $1{\cdot}038\times 10^{-8}$ cm.
Molybdenum radiation .	K_α 0·71,	K_β $0{\cdot}63\times 10^{-8}$ cm.	
Selenium radiation	K_α 1·104,	K_β $0{\cdot}984\times 10^{-8}$ cm.	

The intensity ratios of the K_α lines and the K_β lines are of the order of 6 : 1. It may be noted that the K_β line of selenium falls between the L_2 and the L_3 absorption edge of gold.

The critical exposures contained in Table I. are the actual measured doses, that is to say the doses falling on the surface of the sol. The average dose received by the sol is less than this owing to the fact that the deeper portions of the liquid are partly screened from the radiation by the absorptive effect of the layers above them. The thickness of the exposed samples was kept as small as possible (0·21 cm. in all cases), but the effect is appreciable for radiations as absorbable as those employed in the present experiments. Since the coefficients of absorption of the two radiations are known the necessary corrections can be made. If μ is the mass coefficient of absorption of the radiation in the sol, ρ the density and d the thickness, the fraction of the incident energy absorbed in the sol is $1-\epsilon^{-\mu\rho d}$. In the absence of any screening it would be $\mu\rho d$. The measured dose must therefore be reduced in the ratio $(1-\epsilon^{-\mu\rho d})/\mu\rho d$. Owing to the very small gold content of the sols the absorption in the gold particles did not exceed 0·3 per cent. and the sol could simply be regarded as water.

The ratio of E_s, the energy absorbed per unit volume of sol, to that absorbed per unit volume of air in the ionization chamber for the same beam is given by $\tau_s\rho_s/\tau_a\rho_a$ where τ_s and τ_a are the true coefficients of energy absorption in the sol and in air respectively. ($\tau=\mu-\sigma$, where σ is the mass coefficient of scattering.) The number of röntgen registered by the chamber is directly proportional to the energy absorption in air, since the energy, z, required to form a pair of ions in air is independent of the wave-length of the radiation. Thus

$$E_s/CD=\tau_s\rho_s/\tau_a\rho_a,$$

where D is the dose in röntgen, and C is a constant the value of which is clearly z/e, e being the electronic charge in e.s.u. Thus if E_s is the energy absorbed per unit volume of sol from the molybdenum radiation, E_s' the corresponding value for the selenium radiation,

$$E_s/E_s'=D/D'\times\left[\frac{\tau_s/\tau_a}{\tau_s'/\tau_a'}\right],$$

where D and D′ are the corresponding doses in röntgen, and τ and τ' refer to the coefficients of absorption for molybdenum and selenium radiation respectively. The quantity in the bracket does not differ by more than a few per cent. from unity, when both the wave-lengths concerned are considerably shorter than that of the K absorption edges of the substances concerned. For the two radiations employed in these experiments the most reliable data give the value of 0·96.

In spite of much excellent and painstaking work, X-ray absorption coefficients do not appear to be known with the degree of precision which is to be desired. In the following numerical calculations we have used values deduced from the extensive table quoted on the authority of S. M. J. Allen by Compton and Allison (6). Fortunately the result is not greatly influenced by the values assumed for the coefficients of absorption, as the correcting factor for absorption in the finite depth of the sol only varies slowly with μ, and the ratio $(\tau_s/\tau_a)/(\tau_s'/\tau_a')$ cannot on theoretical grounds differ appreciably from unity.

We have already seen that for each type of radiation D_n/k^n is a constant. The ratio E_s/E_s' of the energies absorbed from the beam in producing a given effect in the sol can thus be best summed up by a comparison of these constants. The values used and obtained in the calculations are given in Table IV.

TABLE IV.

Radiation.	μ_s.	$\frac{1-\epsilon^{-\mu\rho d}}{\mu\rho d}$.	D_n/k^n obs.	D_n/k^n corr.
Mo_K	1·09	0·90	1·36	1·22
Se_K.........	3·57	0·70	1·05	0·74

$$E_s/E_s'=(1\cdot22/0\cdot74)\times\frac{\tau_s\tau/a}{\tau_s'/\tau_a'}=1\cdot65\times0\cdot96=1\cdot58.$$

Calculation shows that, in order to produce a given change in the electrophoretic mobility of the sol, the energy absorbed per unit volume from the molybdenum radiation is 1·58 times that absorbed from the selenium radiation.

No correction has been made for the obliquity of the peripheral radiation, which would increase the effective value of the thickness of the sol. However, this correction would not be sufficiently large to make an appreciable alteration in the value of the ratio of the energies absorbed in the sol. Assuming that $z=33$ volts per ion pair, and $e=4\cdot80\times10^{-10}$ e.s.u., the energy absorbed in the sol for Mo_K radiation works out at 80 erg/cm.^{3} per röntgen,

The mean wave-length of Mo_K radiation is $0\cdot70\times10^{-8}$ cm., and the quantum energy is therefore $2\cdot8\times10^{-8}$ erg. Taking the value of 80 erg per c.c. per röntgen as approximately correct, the number of photons absorbed per c.c., when a dose of one röntgen has been received, is $80/2\cdot8\times10^{-8}=2\cdot8\times10^{9}$. The number of gold particles per c.c. was $5\cdot8\times10^{9}$. The number of photons absorbed in the sol per particle during a dose of one röntgen is thus only 0·5. A dose of 2 röntgen is sufficient to reduce the mobility of the sol to the first minimum. This corresponds to an absorption in the whole liquid of only about one photon per particle. As has already been pointed out, only a small fraction of the absorption in the sol takes place in the gold particles themselves. It seems impossible, therefore, that the effect can be due to a direct hit scored by the photons on the colloid particle.

On the other hand, it is interesting to note that the ratio $E_s/E_s'=1\cdot58$ is almost exactly equal to $\lambda'/\lambda=1\cdot57$, where λ is the mean wave-length of Mo_K and λ' that of Se_K radiation. The number of quanta absorbed in the producing of a given change in the electrophoretic mobility of the sol is thus the same for both types of radiation. Experiments with other wave-lengths would be necessary to test whether this is anything more than a curious coincidence. The agreement is certainly surprisingly close. Owing to war conditions it seems improbable that these experiments will be possible in the near future. We place them on record for further investigation when circumstances permit.

Summary.

Experiments with a series of gold sols of different particle size which were irradiated with gamma radiation and homogeneous X-radiation of two different wave-lengths have given the following results :—

(1) The electrophoretic mobility of the sols shows rhythmic increases and decreases with steadily increasing doses of radiation.

(2) If the D_n is the dose required to bring the sol to the *n*th turning-point (maximum or minimum) of the curve relating dose with electrophoretic mobility, for a given sol the relation holds :

$$D_n/k^n=\text{constant}.$$

The value of k is independent of the particle size, the nature of the sol and the radiation employed within the limits of experimental error.

(3) The quantity a^2D_n/k^n is approximately constant for sols of the same material, but decreases slowly with increase in the radius a.

(4) The energy absorbed per unit volume of sol for the production of a given change in the electrophoretic mobility of the sol is 1·58 times greater for Mo_K radiation than for Se_K radiation. For these radiations the number of quanta absorbed for a given change in the mobility is the same. There is not sufficient evidence to show whether this is a general rule for all wave-lengths.

The work recorded in this paper has been made possible by a grant from the British Empire Cancer Campaign, to whom we would express our grateful thanks. One of us (R. J.) acknowledges with thanks the assistance of a training grant from the Department of Scientific and Industrial Research.

References.

(1 *a*). Crowther, Liebmann, and Lane, Phil. Mag. ser. 7, xxiv. p. 654 (1937).
(1 *b*). Crowther, Liebmann, and Jones, Phil. Mag. ser. 7, xxvi. p. 120 (1938).
(1 *c*). Crowther, Liebmann, and Jones, Phil. Mag. ser. 7, xxviii. p. 64 (1939).
(2) Crowther, 'Nature,' cxlii. p. 569 (1938).
(3) Westgren, *Z. anorg. Chem.* xciii. p. 151 (1915).
(4) Pennycuik and Woolcock, 'Australian Journal of Experimental Biology,' xv. p. 423 (1937).
(5) Rymer, 'Journal of Scientific Instruments,' xvi. p. 84 (1938).
(6) Compton and Allison, 'X-rays in Theory and Experiment,' Macmillan & Co., London: Second Edition, 1935.

The University of Reading.
December 1939.

XL. *Observations on Edge Tones.*

By J. M. A. Lenihan, B.Sc., and E. G. Richardson, B.A., Ph.D., D.Sc., (University of Durham, King's College, Newcastle-upon-Tyne) *.

[Received July 15, 1939.]

Abstract.

Experiments are described on the tones which are heard when air issuing from a slit strikes a parallel wire. The tones are shown to be of the same nature as edge tones. In the velocity gradient on either side of the slit the wire may be set into vibration at its natural frequency. The significance of the observations for edge-tone theories is discussed.

The problem of edge tones is one which continues to form a battle-ground for rival theories, though a complete solution seems as far off as ever. The object of this paper is to report further experiments which, while not by any means saying the last word on this matter, may throw some more light on it, dealing as they do with a new class of edge tone.

An edge tone is produced whenever a jet of fluid strikes a wedge placed in its path, as in the flute and diapason organ pipes. A complete bibliography of the phenomenon will be found in the recent papers of Burniston Brown †. In the discussion following the reading of these papers it was suggested by one of us (E. G. R.) that information on the mechanism of an edge tone might be obtained by replacing the usual wedge with a thin wire, since in this case the building up of pressure excesses on either side of the wedge, which is an essential feature of the " escapement " theory mentioned later, might be made difficult. Since this suggestion was put forward, Carrière ‡ and Coop § have described independent experiments on the sounds produced when a wire is held in an *ignited* jet. The need for some clarification is shown by the fact that, whereas Carrière finds a ruling parameter involving the distance of the burner orifice from the wire, and therefore of an edge-tone type, Coop finds a frequency dependent principally on the *diameter* of the wire, and classes it among æolian tones.

---

* Communicated by the Authors.

† G. Burniston Brown, Proc. Phys. Soc. xlix. pp. 493 and 508 (1937); 'Nature,' cxli. p. 11 (1938); 'Science Progress,' xxxiii. p. 129 (1938).

‡ Z. Carrière, *Rev. d'Acoust.* v. p. 112 (1936).

§ J. J. Coop, J. Acoust. Soc. Amer. ix. p. 321 (1938).

In our opinion, the ignited jet introduces an unnecessary complication into the experiment in the form of uncertain temperature and velocity fields. Another uncertainty was introduced into the experiments mentioned above by using the mean velocity of the jet, calculated from the supply pressure; it is, of course, preferable to measure the velocity directly in the vicinity of the wire. As the wire tone is produced in a region closer to the orifice than a wedge tone, accurate velocity measurements are especially important.

The wind supply for our experiments was obtained from an electric blower, passed through a bellows to stabilise the pressure, and then into a wooden box, 4′ long and 15″ square in section, tapered to 12″ square at the exit end. The slit was a metal one with adjustable jaws, fixed to one end of the wind chest. To make the exit of the wind as smooth as possible, the corners of the box were packed inside with plasticene, and the jaws of the slit were bevelled. The convergence of the stream commenced about 12″ before the exit. The slit was vertical and about 2″ long, delivering a stream of air at the wire, which was held in a hacksaw type of frame. This frame was mounted on a lathe bed which could be moved both longitudinally and transversely with respect to the slit. The velocity field was explored with a Pitot tube, formed of hypodermic metal tubing, connected to a pressure gauge. For pressures up to 2 cm. of water a Chattock tilting gauge, reading to 0·002 cm., was used; higher pressures were measured by a water manometer and a travelling microscope, with an accuracy of 0·005 cm. The note produced was picked up by a ribbon microphone held near the wire, and the microphone current, after amplification, was fed into a harmonic wave analyser. The response of the analysing system was previously checked with a beat-frequency oscillator of calibrated output. The diameter of the wire was 0·47 mm., and the width of the slit was 1·9 mm.

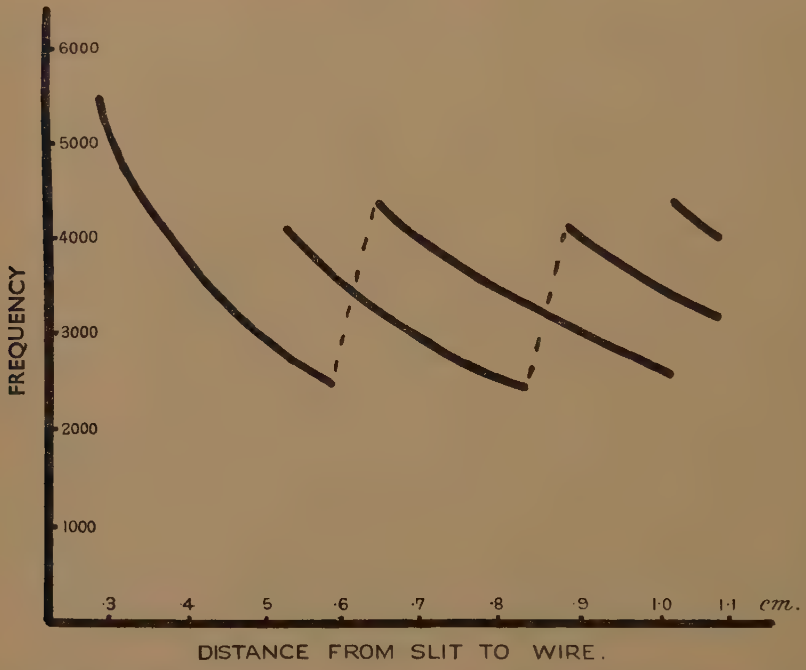
The first work undertaken was a determination of the edge-tone constant, V/af (V=velocity of stream, a=distance from slit to wire, f=frequency of note). Measurements of the frequency and velocity were made at different points in the stream. A typical set of results for frequency against distance is shown in fig. 1. It will be seen that the note exhibits the frequency jumps familiar to edge tones. The note usually contained two components, with frequencies in the ratio 1·4 : 1 approximately; this is accounted for below. The edge-tone constant V/af is reasonably constant for each set of observations, though the results do not agree with the simple formula $V/af=j/2$(j integral), which is commonly applied to edge tones. The æolian tone constant, V/fd (d=diameter of wire), varied considerably even for one wire. By mounting a set of wires varying in diameter from 2 to 0·2 mm. on the same frame, so that they could be moved in turn to the same point in the stream, it was found that the diameter, within these limits, produced no effect on the frequency of the tone; with thin wires, however, the higher harmonics were more prominent.

The formula

$$f = \cdot 466j(V-40)(1/a-0\cdot 07),$$
$$j = 1,\ 2\cdot 3,\ 3\cdot 8,\ 5\cdot 4,\ \text{etc.}$$

has been proposed for edge tones by Brown *. In this case V is the mean velocity, calculated from the supply pressure; as in our experiments the velocity was measured directly, it is not necessary to introduce the figure 40, which presumably corrects for the falling-off in velocity at a distance from the slit. To test Brown's formula, the frequency of the note was calculated in each case, using the amended form:

Fig. 1.



$$f = \cdot 466jV(1/a-0\cdot 07).$$

A suitable value of j was found by approximate calculation. We add two values of j, 7·2 and 9·2, to the series given by Brown, and we find in one case $j=2\cdot 7$. The calculated and observed frequencies are set out in the table, and exhibit quite striking agreement. Brown claimed an

---

* *Loc. cit.*

accuracy of only 6 per cent. for his formula; the amended form fits our results with a greater accuracy than this.

TABLE I.

Diameter of wire=·47 mm. Width of slit=1·9 mm.
(a is in cm.; V in cm./sec.)

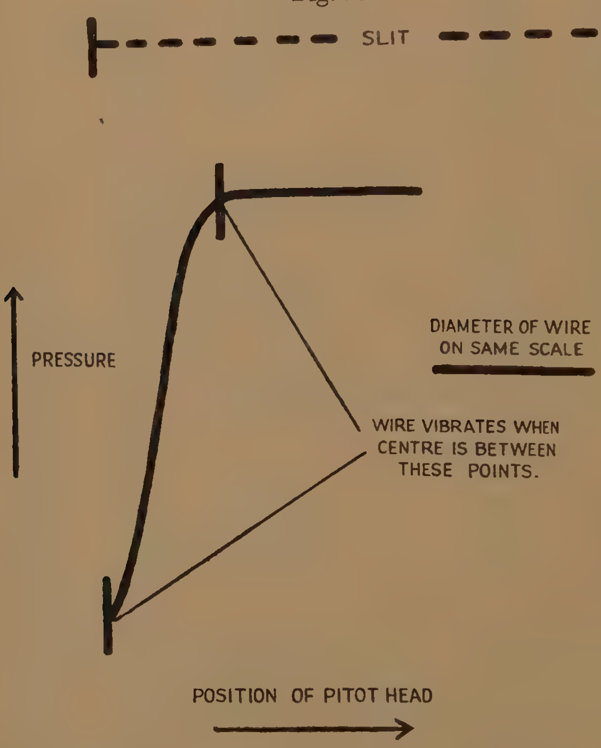
a.	V.	j.	f calc.	f obs.	V/af.
·304	1340	2·7	5430	5440	·810
·364	1325		4460	4480	·812
·423	1308		3775	3830	·807
·483	1296		3260	3240	·828
·542	1278		2855	2820	·836
·602	1268		2540	2550	·826
·542	1278	3·8	4010	4140	·570
·602	1268		3575	3700	·570
·661	1237		3145	3280	·570
·720	1222		2855	2950	·574
·780	1198		2570	2670	·575
·839	1160		2035	2460	·562
·546	995	5·4	4400	4450	·416
·576	971		4090	4060	·415
·605	949		3775	3840	·408
·634	914		3470	3600	·400
·898	1153	7·2	4090	4100	·316
·958	1127		3780	3720	·314
1·017	1106		3400	3440	·331
1·077	1090		3135	3200	·316
1·017	1106	9·2	4340	4400	·247
1·077	1090		4160	4000	·250
·661	1237	5·4	4480	4370	·428
·720	1222		4065	3900	·435
·780	1198		3650	3610	·425
·839	1175		3310	3360	·418
·898	1153		3035	3100	·416
·958	1127		2825	2820	·411
1·017	1106		2550	2640	·416
1·077	1090		2350	2440	

Vibrations of the Wire.

The wire could be made to vibrate at its fundamental frequency by placing it in a region where there was a steep gradient of velocity. The largest velocity gradient occurred at the sides of the jet (see fig. 2), and here the wire vibrated in a plane perpendicular to the direction of flow of the stream. Near the slit there was a considerable velocity gradient parallel to the stream, and the wire vibrated here also ; in the appropriate position these two vibrations were executed simultaneously as a Lissajous figure.

Observations on this phenomenon were made using a slit with fairly blunt jaws ; with sharp jaws, the velocity rises very steeply at the sides of the stream, and the wire vibrates in a very small region. The vibration appears to be a relaxation phenomenon. The difference of pressure on the two sides of the wire causes it to be pushed into the low-pressure region (Magnus effect) until components of tension cause it to spring back. On recovery it is again urged inwards, and so on. In the central region of nearly uniform velocity, the wire did not appear to vibrate, even when the wind-tone was in resonance with the natural frequency of the wire.

Fig. 2.


SLIT
PRESSURE
DIAMETER OF WIRE ON SAME SCALE
WIRE VIBRATES WHEN CENTRE IS BETWEEN THESE POINTS.
POSITION OF PITOT HEAD

The pressure gradient and amplitude of the vibration were measured, and it was found that the amplitude for a given wire was

(i.) proportional to the pressure gradient ;

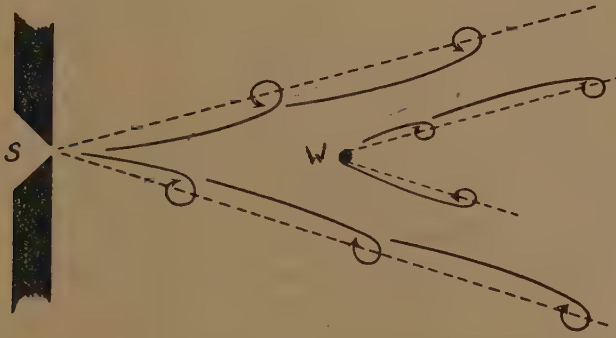
(ii.) inversely proportional to the square of the natural frequency of the wire. This, of course, involves the tension, mass and length of the wire.

Theories of the Phenomenon.

In order to discuss the significance of these experiments for edge tones, we must outline the various theories which have been put forward to explain this vexed problem.

1. *The "Escapement" Theory.*—Originally put forward by Wachsmuth *, this has been revived in a rather different form by Burniston Brown. A slight deflexion of the jet to one side of the wedge builds up an excess pressure on that side, which causes the succeeding fluid to deviate to the other side, where the pressure again builds up.

Fig. 3.



2. *The "Shock-wave" Theory.*—A series of vortices leaving the orifice travels along the jet. When each one hits the opposing edge, a pressure wave travels with the velocity of sound back to the orifice, where it starts another one practically instantaneously (Krüger †).

3. The pendulation of the jet is the result of two superposed systems of Bénard-Kármán vortices, one set originating at the slit and the other at the edge (Richardson ‡). In the present instance the latter set are clearly formed behind the wire, after the fashion of æolian tones. From the wire on there are then four lines of vortices instead of two (*cf.* fig. 3). The only way in which the procession can marshal itself into a Bénard-Kármán double avenue of vortices is for one to roll from the wire (W) at the same time as another rolls up from the slit (S) itself.

The fundamental observation which has to be explained by the theory is, of course, that the pendulation or vortex motion which gives rise to the sound is periodic with a wave-length bearing a constant integral or fractional relationship to the slit-edge distance. Since an edge tone can still be obtained when the wedge is reduced to a fine wire, it seems difficult to conceive how the pressure difference required by (1) can be set up in such a case. Furthermore, (1) does not lead to a simple explanation of Brown's formula, though (2) does agree with the simple law $V/af=j/2$. A crucial experiment was carried out by one of us (E. G. R.) * a few years ago to test the shock-wave theory. In an edge-tone apparatus of the type in which a coloured stream of *liquid* strikes an edge immersed in a bath of the same liquid a second slit was opened in the supply chest, 5 cm. to the right of the existing one, but not provided with an opposing wedge. The first jet having been set into rhythmic pulsation by its opposition to the edge, the postulated compression sent out by the embryo vortices striking the wedge should have travelled in all directions through the medium and set the second free jet into pulsation at the same rate as the first ; actually, no pendulation of this second jet was to be seen. Incidentally, Krüger † does not consider this an objection to his theory, as he opines that the reflected pulse which he postulates would be *confined to the jet* which impinges on the edge. This does not seem a valid defence in our view, since the jet is composed of the same material as its surroundings, and, moreover, is running at a speed little faster than the circumambient fluid, so that there is no sharp surface of discontinuity along the jet, which could confine the shock wave to the stream in which it arises, as in the case of jets moving at supersonic speeds (*cf.* the Hartmann oscillator).

---

* R. Wachsmuth, *Ann. Phys., Lpz.* xiv. p. 469 (1904).

† F. Krüger, *Ann. Phys., Lpz.* lxii. p. 673 (1920).

‡ E. G. Richardson, Proc. Phys. Soc. xliii. p. 394 (1931).

Our evidence, rather negative in nature, therefore brings us back to (3), which, like (1), is a purely hydrodynamic theory, *i. e.*, it does not demand the existence of density changes in the fluid. The fact that a coloured jet of water can be made to pendulate if it strikes an edge, but that the pendulation is inhibited if the jet is constrained on either side by the too close proximity of glass plates (as in the Hele Shaw apparatus), suggests that the primary phenomenon should be capable of mathematical analysis in terms of the flow of a fluid of small viscosity, with neglect of compressibility, as in many allied phenomena of aerodynamics, and that the tone itself is a secondary phenomenon, in which compressions arise due to the vortices striking the obstacle or actually setting it into vibration. The form of that mathematical analysis is, however, as elusive to us as it has proved to our predecessors and companions in this pursuit. The object of this note is merely to weigh the evidence which has come forward during the last few years.

---

* *Loc. cit.*

† F. Krüger and H. Casper, *Phys. Z.* xxxvii. p. 842 (1936).

XLI. *Waves produced by the Rolling of a Ship.*

By T. H. Havelock, F.R.S.*

[Received December 21, 1939.]

1. The first part of the following paper deals with the surface waves produced by an elliptic cylinder, or a flat plate, submerged in water and performing small linear or rotational oscillations. The second part contains a short discussion of the energy dissipated in wave motion by a rolling ship, together with an estimate of the magnitude of this effect obtained from the preceding results.

Submerged Elliptic Cylinder.

2. The method adopted is to replace the oscillating body by some suitable distribution of sources and sinks or of doublets. Although the analysis could be extended to three-dimensional problems, we limit consideration at present to two-dimensional motion in a frictionless liquid. We begin with the solution for a horizontal doublet which was obtained for an oscillating circular cylinder (Havelock, 1917). Take the origin O in the free surface of deep water, Ox horizontal and Oy vertically upwards. Let there be a horizontal doublet of oscillating moment M cos σt at the point $(0, -f)$ in the liquid. The velocity potential ϕ is given by

$$\phi=\frac{\mathrm{M}x}{r_1^2}e^{i\sigma t}+\mathrm{M}e^{i\sigma t}\int_0^\infty\frac{g\kappa+\sigma^2-i\mu'\sigma}{g\kappa-\sigma^2+i\mu'\sigma}e^{-\kappa(f-y)}\sin\kappa x\,d\kappa, \quad . \quad . \quad (1)$$

with $r_1^2=x^2+(y+f)^2$. The real part of the expression is to be taken, and, further, the limiting value when the positive quantity μ' tends to zero; this latter process ensures that, at great distances from the origin, the motion will reduce to waves travelling outwards on either side. We may write (1) in the form

$$\phi=\frac{\mathrm{M}x}{r_1^2}e^{i\sigma t}+\frac{\mathrm{M}x}{r_2^2}e^{i\sigma t}+2\kappa_0\mathrm{M}e^{i\sigma t}\int_0^\infty\frac{e^{-\kappa(f-y)}\sin\kappa x}{\kappa-\kappa_0+i\mu}d\kappa, \quad . \quad . \quad . \quad (2)$$

with $r_2^2=x^2+(f-y)^2$ and $\mu=\mu'\sigma/g$.

The integral in (2) may be transformed by taking κ to be a complex variable and integrating round suitable contours according as x is positive or negative. We obtain, taking the real part and making μ zero,

---

* Communicated by the Author.

$$\phi=\frac{\mathrm{M}x}{r_1^2}\cos\sigma t+\frac{\mathrm{M}x}{r_2^2}\cos\sigma t\mp 2\kappa_0\mathrm{M}e^{-\kappa_0(f-y)}\cos(\sigma t\mp\kappa_0 x)$$

$$\mp 2\kappa_0\mathrm{M}\cos\sigma t\int_0^\infty\frac{\kappa_0\cos\kappa(f-y)+\kappa\sin\kappa(f-y)}{\kappa^2+\kappa_0^2}e^{\mp\kappa x}d\kappa, \quad (3)$$

the upper or lower signs to be taken according as x is positive or negative.

The corresponding surface elevation η is given by

$$\eta=\mp\frac{2\pi\kappa_0\sigma\mathrm{M}}{g}e^{-\kappa_0 f}\sin(\sigma t\mp\kappa_0 x)-\frac{2\mathrm{M}\sigma}{g}\frac{x}{x^2+f^2}\sin\sigma t$$

$$\pm\frac{2\kappa_0\mathrm{M}\sigma}{g}\sin\sigma t\int_0^\infty\frac{\kappa_0\cos\kappa f+\kappa\sin\kappa f}{\kappa^2+\kappa_0^2}e^{\mp\kappa x}d\kappa. \quad (4)$$

The first term represents the regular waves, while the other two terms give a local oscillation whose magnitude diminishes with increasing distance from the centre of disturbance.

Similar expressions may be obtained for a source of oscillating magnitude, or for a doublet with its axis in any direction. It may be remarked that for a doublet at a given point in the liquid, so far as the regular waves are concerned the direction of the axis affects only the phase of the waves and not their amplitude.

3. Consider the motion produced by an elliptic cylinder moving through an infinite liquid. If the motion of the cylinder is one of translation, it is well known that the fluid motion is that due to a certain distribution of doublets along the line joining the foci of the elliptic section of the cylinder; a similar proposition may also be readily proved when the motion is one of rotation.

In particular, let the cylinder be moving with velocity V parallel to the minor axis of the section; let S, S′ be the foci of the section and h the distance of a point on SS′ from the centre C. The moment per unit length of the doublet distribution is

$$a\mathrm{V}(a^2e^2-h^2)^{\frac{1}{2}}/\pi(a-b), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

in the usual notation, the axes of the doublets being perpendicular to SS′.

If the cylinder is rotating round C with angular velocity ω, the moment per unit length along SS′ is

$$\omega(a+b)h(a^2e^2-h^2)^{\frac{1}{2}}/2\pi(a-b), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

the axes being perpendicular to SS′.

Combining (5) and (6) with a suitable value of V, we may obtain the distribution when the cylinder is rotating about any point on the major axis.

4. Suppose that the cylinder is wholly immersed in liquid with the axis of the cylinder horizontal and at a depth f below the free surface of the liquid, and let the cylinder be making small rotational oscillations about its axis. Let the angle θ between the major axis of the section and the vertical be given by

$$\theta=\theta_0 \sin \sigma t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where θ_0 is small.

For a first approximation we neglect the effect of the free surface. The angular velocity of the cylinder is $\sigma\theta_0 \cos \sigma t$, and the velocity potential is that due to a certain distribution of doublets along the instantaneous position of the major axis. We shall make a further approximation for small oscillations and assume that this distribution is along the mean position of the major axis, that is, the vertical through the centre of the ellipse. Thus we consider the motion to be due to a distribution of horizontal doublets of oscillating magnitude along the line between the foci of the ellipse in its mean position. From (6), the moment per unit length at a distance h from the centre of the ellipse is

$$\frac{\sigma\theta_0(a+b)}{2\pi(a-b)} h(a^2e^2-h^2)^{\frac{1}{2}} \cos \sigma t, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

the limits for h being $\pm ae$.

We replace M in (3) by this expression, write $f+h$ for f, and integrate with respect to h; we obtain then the velocity potential for the given distribution when the condition at the free surface is satisfied. Similarly, from (4) we may obtain complete expressions for the corresponding surface elevation. This consists of a local oscillation whose amplitude diminishes rapidly with distance from the cylinder, together with regular waves travelling out on either side. We shall examine here only the amplitude of these regular waves; from (4) and (8) the amplitude A of these waves, that is, the coefficient of $\sin(\sigma t-\kappa_0 x)$ for positive values of x, is given by

$$\begin{aligned} A &= -\kappa_0{}^2\theta_0 \frac{a+b}{a-b} \int_{-ae}^{ae} h(a^2e^2-h^2)^{\frac{1}{2}} e^{-\kappa_0(f+h)} dh \\ &= -\kappa_0{}^2(a+b)^2\theta_0 e^{-\kappa_0 f} \int_0^{\pi} \sin^2\theta \cos\theta \, e^{-\kappa_0 ae \cos\theta} d\theta. \quad . \quad . \quad . \quad (9) \end{aligned}$$

This may be expressed in terms of the modified Bessel function $I_n(x)$, and we obtain

$$A=\frac{\pi\theta_0(a+b)^{\frac{3}{2}}}{(a-b)^{\frac{1}{2}}} \{\kappa_0 ae I_0(\kappa_0 ae)-2I_1(\kappa_0 ae)\} e^{-\kappa_0 f}. \quad . \quad . \quad . \quad (10)$$

If $\kappa_0 ae$ is small, that is, if the wave-length is large compared with the linear dimensions of the cylinder, the first term in the expansion of (10) gives, as an approximation,

$$A=\tfrac{1}{8}\pi(a-b)(a+b)^3\kappa_0{}^3\theta_0 e^{-\kappa_0 f}. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Consider a cylinder of given vertical dimension $2a$, with varying breadth $2b$. Naturally, for the circular cylinder $(b=a)$ the disturbance is zero. It is of interest to note that for the approximation (11) the maximum wave amplitude occurs for $b=\frac{1}{2}a$, its value being then nearly twice the value for the limiting case of the flat plate $(b=0)$.

Suppose that the cylinder has its major axis vertical, and is making linear horizontal oscillations in which the displacement is $d\sin\sigma t$. Then from (4) and (5) the amplitude of the regular waves is

$$A=\frac{2\kappa_0{}^2ad}{a-b}\int_{-ea}^{ae}(a^2e^2-h^2)^{\frac{1}{2}}e^{-\kappa_0(f+h)}dh$$

$$=2\pi\kappa_0 ad\left(\frac{a+b}{a-b}\right)^{\frac{1}{2}}I_1(\kappa_0 ae)e^{-\kappa_0 f}. \tag{12}$$

If $\kappa_0 ae$ is small, this gives, approximately,

$$A=\pi\kappa_0^2a(a+b)de^{-\kappa_0 f}. \tag{13}$$

Finally, combining (10) and (12) with $d=a\theta_0$, we obtain the amplitude for an elliptic cylinder with its major axis vertical in its mean position, and making small angular oscillations given by $\theta=\theta_0\sin\sigma t$ about the upper end of its major axis ; in this case we obtain

$$A=\pi\theta_0\left(\frac{a+b}{a-b}\right)^{\frac{1}{2}}\{\kappa_0 ae(a+b)I_0(\kappa_0 ae)-2(a+b+\kappa_0 a^2)I_1(\kappa_0 ae)\}e^{-\kappa_0 f}. \tag{14}$$

For $\kappa_0 ae$ small, the first term in the expansion is the same as (13) with $d=a\theta_0$.

In all these cases the expressions take simpler forms in the limiting case of the flat plate, for which we made b zero ; but it should be noted that the ideal solution then implies infinite fluid velocity at the edges of the plate. In particular, consider a plate of height $2a$, making small oscillations about its upper edge, the centre of the plate being at a depth f. If $\kappa_0 a$ is small, the first term in the expansion of (14) gives

$$A=\pi\kappa_0{}^2a^3\theta_0e^{-\kappa_0 f}. \tag{15}$$

This, naturally, is equivalent to replacing the oscillating plate by a single doublet at its centre. If, in addition, $\kappa_0 f$ is small, we may take $\pi\kappa_0{}^2a^3\theta_0$ as a first approximation for the amplitude of the regular waves. A similar approximation could be made for a cylinder of any cross-section, using the corresponding inertia coefficient for linear motion and the mean horizontal velocity of the cylinder.

Rolling Ship.

5. The expressions given in the previous section are approximations suitable for wholly submerged bodies ; it is not permissible, in general, to apply them to the oscillations of floating bodies. The approximation

used for the doublet distribution loses accuracy with diminishing depth of submergence of the body; moreover, when the surface of the body cuts the free surface some of the expressions for the surface elevation may take infinite values. It may be noted, however, that such infinities generally occur in the local part of the disturbance, the expressions for the amplitude of the regular waves at a distance from the body remaining finite.

For the rolling ship the period is such that the wave-length of the corresponding waves is large compared with the draught of the ship. Thus if we consider the analogous problem of the oscillating plate with its upper edge in the surface, the quantity $\kappa_0 a$ of the previous section is small, in most cases about 0·1. In these circumstances, treating the motion as two-dimensional, we propose to regard the ship as a single oscillating doublet at a depth which is small compared with the wave-length; thus, from (4), we take $2\pi\kappa_0\sigma M/g$ as a first approximation for the amplitude of the waves at a distance from the ship. Further, as we cannot expect more than an estimate of the order of magnitude from this assumption, we shall regard the ship as a plank, of length L and draft D, oscillating about the water-line through an angle θ_0 on either side of the vertical; using the result given at the end of § 4 and writing T for the complete period of rolling, this gives for the height of the regular waves

$$h=\frac{4\pi^5 D^3}{g^2 T^4}\theta_0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

It should be noted that the wave-height, as the term is commonly used, is measured from trough to crest and is twice the amplitude.

6. Before applying this result, we may review briefly calculations which have been made from a different point of view.

The part played by wave propagation in causing resistance to rolling was first recognized by W. Froude (1872) and was advocated by him in a series of papers. Froude showed that the energy propagated outwards in the wave motion corresponds to a resisting couple proportional to the angular velocity of rolling, and also that the energy actually dissipated in rolling, or a large part of it, could be accounted for by waves of extremely small height; in one case, for example, his calculation gave a height of $1\frac{1}{4}$ inches for waves 320 ft. long. The same method has been applied by other writers subsequently, and it may be worth while repeating the argument in a somewhat different form from that in which it is usually given.

Suppose the ship to be rolling about a horizontal axis through its centre of gravity, and take the equation of motion in its simplest form as

$$I\ddot{\theta}+N\dot{\theta}+Wm\theta=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where I is the moment of inertia, W the weight, m the metacentric height, and $N\dot{\theta}$ the resisting couple.

The exact solution of (17) gives damped oscillations with a damping coefficient $h=N/2I$, and the rate of dissipation of energy is $N\dot{\theta}^2$. Suppose now that the dissipation is small and assume an undamped motion $\theta=\theta_0 \sin \sigma t$, holding approximately for a sufficient time, with

$$\sigma^2=Wm/I=4\pi^2/T^2.$$

With this assumption, the average rate of dissipation of energy

$$=\tfrac{1}{2}N\sigma^2\theta_0{}^2=Wmh\theta_0{}^2.$$

In the usual notation for the rolling of ships, $a\theta_0$ is the decrement of rolling angle for one swing ; hence $a=\frac{1}{2}hT$. Thus the average rate of dissipation of energy is $2Wma\theta_0{}^2/T$. Assume, with Froude, that when the ship is rolling, regular straight-crested waves are sent out on either side, the breadth of each train being approximately equal to the length L of the ship ; further, let A be the amplitude of the waves, λ the wave-length, T the period, with $\lambda=gT^2/2\pi$. In each train energy is propagated outwards at half the wave velocity V, that is, at an average rate $\frac{1}{4}g\rho A^2VL$ on each side. Hence, equating the average rate of dissipation of energy to the average rate of propagation of energy outwards in the waves on both sides, we have

$$2Wma\theta_0{}^2/T=\tfrac{1}{2}g\rho A^2V\lambda,$$

or

$$Wma\theta_0{}^2=\tfrac{1}{4}g\rho A^2\lambda L. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

This is, in effect, the equation given by Froude and used by later writers, the left-hand side of (18) being the loss of energy in one swing ; the other side of Froude's equation was, however, twice that given in (18), owing apparently to neglect of the difference between group velocity and wave velocity. The statement given here, besides including this correction, shows the various assumptions and brings the argument into line with the usual method of approximating to the damping coefficient in isochronous damped oscillations when the damping is sufficiently small. Froude recognized that his solution was not in any sense rigorous and hoped that it would be supplemented by some direct estimate, even if with no greater exactness, of the wave-making property of a ship when rolling ; it is also of interest that he proposed to attempt direct observation of the waves produced by the rolling of models of simple form. However, nothing further seems to have been done on this particular aspect of the problem since that time. Other writers have used Froude's expressions to estimate the wave height, and it appears to be accepted that wave motion accounts for a large part of the dissipation of energy in rolling, that due to fluid friction or eddy-making being relatively small apart from exceptional resistance due to bilge keels; it has been remarked, for instance, that no reasonable values of head resistance and skin friction coefficient account for more than one-third of the actual decrement obtained by experiment, and in one case such a calculation gave only one-seventeenth of it (Baker, 1914). Nevertheless, no attempt appears to have been made to compute the wave resistance to rolling from the characteristics of the ship.

7. We shall now compare wave heights calculated from (16) with various cases to which Froude's energy method has been applied.

In the case examined in Froude's first paper already quoted, the data are $T=8$ sec.; $\theta_0=5\cdot65°$. The draught of the ship was not stated, but we may assume $D=15$ ft. With these values, (16) gives $h=1\cdot2$ inch. Froude's estimate from energy dissipation was a wave height of $1\frac{1}{4}$ inch. Other writers who have used the same formula assume that that part of the resisting couple which is proportional to the angular velocity of rolling may be attributed to the loss of energy in surface-waves. Thus Sir W. White (1895), for the rolling of H.M.S. 'Revenge' without bilge keels, deduced a wave height of about $1\frac{1}{2}$ inch. In this case $T=15\cdot5$ sec.; $\theta_0=13°$; $D=27$ ft.; and these give from (16) a wave height of just over 1 inch.

L. Spears (1898), from the rolling of U.S.S. 'Oregon,' deduced a wave height of 0·62 inch. Here $T=15\cdot2$ sec.; $\theta_0=12°$; $D=23$ ft.; and (16) gives a wave height of 0·67 inch.

It should be remarked that in all these cases Froude's formula was used; according to the argument given in § 6 and expressed in equation (18), these estimates of wave height should be increased by a factor $\sqrt{2}$. A final example is taken from a recent paper by G. S. Baker (1939) on the rolling of ships under way. We take the data for model R 8(*a*), for rolling at zero speed ahead, given in Tables 1 and 3 of the paper; in the notation already used

$$W=10{,}150 \text{ ton}; \quad m=4\cdot4 \text{ ft.}; \quad T=11\cdot52 \text{ sec.};$$

$$\lambda=680 \text{ ft.}; \quad L=400 \text{ ft.}; \quad D=23\cdot2 \text{ ft.}; \quad a=0\cdot022.$$

In this case we shall use equation (18) to see what height of waves would suffice to account for the whole of the dissipation of energy, neglecting for the moment any due to friction or eddy making. With the given values we find from (18), $h=2A=2\cdot65$ inch. Again, using the values of D and T in (16), we find $h=1\cdot58$ inch.

It should be noted that (16) was derived by regarding the ship as a thin plank. The formula could be modified in an empirical manner to take into account the displaced volume and the inertia coefficient of the ship; this might be represented by multiplying (16) by a factor whose probable value would lie between 1 and 2, but the modification is not worth while at this stage.

Both the energy method and the present calculation are no more than first approximations, and therefore we may not attach any great accuracy to the estimates by either method ; nevertheless, it is interesting that both methods give results of the same order of magnitude. On the theoretical side the problem should be treated as three-dimensional, and also the boundary conditions at the surface of the ship satisfied more closely ; in addition, the actual motion of the ship and its axis of rotation are important factors in a more detailed investigation. On the other hand, it would be desirable to have experiments on models of suitable form designed to provide better estimates of frictional and eddy-making resistance to rolling, and so to afford more reliable knowledge of the amount left to be accounted for by wave propagation.

Summary.

Expressions are obtained for the surface disturbance produced by a cylinder, of elliptic cross-section, submerged in water and making small oscillations. A simple form of these results is used as a first approximation for the height of the waves, supposed two-dimensional, sent out on either side by a rolling ship. Numerical calculations are made for cases for which a similar estimate has been made by an energy method due to W. Froude ; the results by the two methods are of the same order of magnitude.

References.

G. S. Baker, Trans. Nav. Arch. lvi. p. 238 (1914) ; also Trans. N.E. Coast Inst. Eng. and Ship. (1939).
W. Froude, ‘Naval Science,’ i. p. 411 (1872) ; also iii. p. 312 (1874).
T. H. Havelock, Proc. Roy. Soc. A, xciii. p. 522 (1917).
L. Spears, Trans. Soc. Nav. Arch. vi. p. 21 (1898).
W. White, Trans. Nav. Arch. xxxvi. p. 145 (1895).

XLII. *Notices respecting New Books.*

Terrestrial Magnetism and Electricity. Edited by J. A. FLEMING. [Pp. xii+794.] (New York and London: McGraw-Hill, 1939. Price 52*s.* 6*d.*)

TERRESTRIAL magnetism and electricity have not been, in recent years, among the more popular subjects of study by physicists generally. This has been in no small degree due to the difficulty of gaining even an initial appreciation of the broader issues and problems embedded in the accumulated masses of diverse observations. This difficulty will be largely removed by the present volume, the latest addition to a series on the Physics of the Earth, prepared as a result of excellent organizing work by the United States National Research Council. The fourteen contributors, eight of them members of the Department of Terrestrial Magnetism at Washington, together provide a survey of the whole field which will be of value both to those engaged in the more specialized investigations and to those who wish to gain a general idea of the present status of the subject and of its outstanding problems.

Chapters dealing primarily with the magnetic side are contributed by Fleming (general survey), Johnston, Fleming, and McComb (instruments), Heiland (magnetic prospecting), McNish (analysis of magnetic changes, causes of the earth's magnetism) and Bartels (solar terrestrial relationships, general problems). The more purely electrical side is discussed by Gish (survey of atmospheric electricity), Torreson (instruments), Rooney (earth currents), Berkner (radio methods of exploring the outer atmosphere), and Schonland (electrical effects of thunder clouds and shower clouds). A comprehensive treatment of the upper atmosphere is given by Hulburt, and Vegard discusses fully the many problems of the aurora polaris. An admirable bibliographical chapter is due to Harradon. This summarizes the sources of publication and the organizations concerned with terrestrial magnetism and electricity, as well as giving a classified bibliography containing over fifteen hundred references. The volume contains a wealth of clearly annotated half tone illustrations, maps, and diagrams.

It would be invidious to pick out any of the chapters for detailed comment, particularly as there is considerable diversity in their character and aim, some constituting almost popular surveys, and others specialized studies. Together they serve the very valuable purpose of providing, in a single volume, accounts of virtually every type of investigation in the field, with discussions of results, and full references. That there is some overlapping is frequently not a disadvantage, as most of the chapters can be read as self-contained wholes, and it is of interest to find full presentations of opposing views, such as those of Hulburt and of Vegard on the origin of aurorae. For a general reader the volume is somewhat formidable, and even those with some initial knowledge of the subject might have welcomed the inclusion of summaries to the various chapters.

In spite of the vast amount of skill and energy which have been devoted, for many years, to accumulating data about terrestrial magnetism and electricity, an understanding even of many of the major effects is still far from having been reached. None, perhaps, would question the desirability of continuing the present extensive routine observational work, but its value is very problematical unless there is some assurance that adequate consideration is to be given to the task of interpreting the significance of the accumulated data. It is to be hoped that this is borne in mind in the allocation of such funds as are available. It has been realized at the Washington Department of Terrestrial Magnetism that the effective investigation of terrestrial magnetism calls for more than routine magnetic observations. It is doubtful, however, whether progress in the solution of the many immediately outstanding problems will be furthered by the erection of a high voltage electrostatic generator for the study of nuclear physics, a description of which forms the astonishing culmination of the introductory survey of terrestrial magnetism in this book.

E. C. S.

A Text-Book of Quantitative Inorganic Analysis Theory and Practice. By ARTHUR I. VOGEL. [Pp. vii+856.] (Longmans, Green & Company. Price 18*s*.)

THIS new work by Dr. Vogel is a companion volume to the author's 'Text-book of Qualitative Analysis,' and aims at meeting the needs of candidates for all examinations involving quantitative inorganic analysis. The book consists of six chapters entitled :—I. The Theoretical Basis of Quantitative Chemical Analysis (pp. 193) ; II. The Experimental Technique of Quantitative Analysis (pp. 84) ; III. Volumetric Analysis (pp. 195) ; IV. Gravimetric Analysis (pp. 197) ; V. Colorimetric Analysis (pp. 62) ; VI. Gas Analysis (pp. 46) ; together with an appendix (pp. 65) containing a large amount of useful analytical data.

The theoretical section, whilst comprehensive in its range of topics, is uneven in quality and is marred by some loose writing and a rather large number of misprints. Thus we are told (p. 3) "It is important to realise that whilst complete ionisation occurs with strong electrolytes, this does not mean that the *effective concentrations* of the ions are the same at all concentrations, . . ." The sections dealing with precipitation and the related topics, co- and post-precipitation, are disappointingly brief and no data are given to indicate the magnitude of the errors which the occurrence of these phenomena may introduce.

In the subsequent chapters, however, the real merits of the book are apparent and a wealth of material is provided, from which a laboratory course suitable for any grade of students may be devised. After a very full treatment of the standard methods of volumetric analysis, excellent accounts are given of the use of ceric sulphate, potassium iodate, potassium bromate, and of a number of other less usual reagents. The gravimetric section is equally generous in its provisions and includes a well-balanced selection of classical and modern methods for a wide range of estimations. Electrolytic determinations and the analysis of complex materials conclude this section.

The last chapters dealing with colorimetric methods and gas analysis maintain the character of the preceding sections in presenting a more than usually detailed account of standard and recent practice.

Experienced analysts will find much of interest in this book, especially as many of the methods have been tried and modified in the author's laboratory. The book is well illustrated throughout and includes a classified bibliography of analytical literature. A reliable index and copious cross-indexing throughout the text add considerably to its value as a work of reference. Both author and publishers are to be congratulated on producing so comprehensive a book at so modest a price.

J. COLVIN.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

XLIII. *The Theory of the Electric Charge and the Quantum Theory.—* Part II.

By H. T. FLINT, Reader in Physics,
University of London, King's College *.

[Received January 15, 1940.]

1. *The Field Theory of the Electric Charge.*

THE analogy with the photon, which is associated with the field equations of the electromagnetic theory, suggests that a set of field equations of a similar form but in a five-dimensional continuum should be introduced as a complement to the theory of the electron which has been discussed. It is first of all necessary to understand the interpretation of the additional components of vectors and tensors which occur when the continuum of four dimensions is replaced by one of five.

It appears that the theories of physics do not recognize a certain group of quantities as forming a vector but prefer to consider the first four contravariant components as a four vector while the fifth covariant component is regarded as a scalar.

In order to lay stress on the difference of representation the five-dimensional vector is denoted by C with components (C^μ), and the corresponding quantity in the four-vector A with components (A^m) in conjunction with the scalar A_0.

The relations between A and C are

$$C^m = A^m \quad (m=1, 2, 3, 4), \qquad (1)$$

and

$$C_5 = \gamma_{5\mu} C^\mu = A_0 \quad (\mu = 1, 2, 3, 4, 5). \qquad (2)$$

Thus

$$A_0 = \alpha\phi_m C^m + C^5,$$

or

$$C^5 = A_0 - \alpha\phi_m A^m. \qquad (3)$$

The case of the line element is an example of this representation with

$$A_0 = d\theta.$$

The corresponding relations between covariant components can be deduced by writing

$$C_\mu = \gamma_{\mu\nu} C^\nu, \quad A_m = g_{mn} A^n.$$

$$C_m = \gamma_{m\nu} C^\nu = \gamma_{mn} C^n + \gamma_{m5} C^5,$$

---

* Communicated by the Author.

$$=g_{mn}A^n+\alpha\phi_m(\alpha\phi_n C^n+C^5),$$

$$=A_m+\alpha\phi_m A_0. \qquad (4)$$

$$C_5=A_0 \text{ by } (2).$$

The relations between the tensor components can be derived from these results for vectors by representing a tensor as a product of vectors. The relations are required for tensors of the second rank, and these will be worked out.

Let a tensor with component $T^{\mu\nu}$ be represented as the product of two vectors F and G.

Thus $$T^{\mu\nu}=F^\mu G^\nu.$$

For the case $$\mu,\ \nu=m,\ n,$$

$$T^{mn}=F^m G^n,$$

and, if A^{mn} denote the component of the corresponding four-dimensional quantity, it follows from (1) that

$$T^{mn}=A^{mn}. \qquad (5)$$

In the case $\nu=5$,

$$T^{m5}=F^m G^5=F^m(A_0-\alpha\phi_n A^n) \text{ by } (3),$$

which can be written in the form

$$T^{m5}=A^m{}_0-\alpha\phi_m A^{mn}. \qquad (6)$$

A_0^m has the property of a four-dimensional vector since it is the product of F^n and A_0.

Similarly, $$T^{5m}=A_0{}^m-\alpha\phi_n A^{nm}. \qquad (7)$$

The component

$$T^{55}=F^5G^5=(B_0-\alpha\phi_m B^m)(A_0-\alpha\phi_n A^n),$$

and this may be written in the form

$$T^{55}=A_{00}-\alpha\phi_m A^m{}_0-\alpha\phi_n A_0{}^n+\alpha^2\phi_m\phi_n A^{mn}. \qquad (8)$$

In the same way the following covariant and mixed tensor components may be verified :

$$\left.\begin{aligned}
T_{mn}&=A_{mn}+\alpha\phi_m A_{0n}+\alpha\phi_n A_{m0}+\alpha^2\phi_m\phi_n A_{00},\\
T_{m5}&=A_{m0}+\alpha\phi_m A_{00},\\
T_{5m}&=A_{0m}+\alpha\phi_m A_{00},\\
T_{55}&=A_{00}.
\end{aligned}\right\} \qquad (9)$$

$$\left.\begin{aligned}
T_m{}^n&=A_m{}^n+\alpha\phi_m A_0{}^n,\\
T_m{}^5&=A_{m0}+\alpha\phi_m A_{00}-\alpha\phi_n A_m{}^n-\alpha^2\phi_n A_0{}^n,\\
T_5{}^m&=A_0{}^m.
\end{aligned}\right\} \qquad (10)$$

The relations between $\gamma_{\mu\nu}$ and g_{mn} are examples of these tensor relations.

In this case $T_{55}=\gamma_{55}=A_{00}$ and has been placed equal to unity for the purpose of the present theory. $A_{0m}=0$ and $A_{mn}=g_{mn}$. Thus by (9)

$$\gamma_{mn}=g_{mn}+\alpha^2\phi_m\phi_n,$$
$$\gamma_{m5}=\gamma_{5m}=\alpha\phi_m,$$

and by (5), (6), (7), and (8)

$$\gamma^{mn}=g^{mn},$$
$$\gamma^{m5}=\gamma^{5m}=-\alpha\phi^m,$$
$$\gamma^{55}=1+\alpha^2\phi_m\phi^m.$$

Another important set of relations is that which exists between the Christoffel bracket expressions in four and five dimensions. The following table gives the relations in which Γ^m_{nl} denotes the four-dimensional expression and Δ^m_{nl} the corresponding five-dimensional quantity (1).

$$\left.\begin{aligned}&\Delta^{\mu}_{55}=0,\quad \Delta^5_{5m}=\tfrac{1}{2}\alpha^2\phi_p H^p{}_m,\\&\Delta^m_{n5}=-\tfrac{1}{2}\alpha H^m{}_n,\\&\Delta^m_{nl}=\Gamma^m_{nl}-\tfrac{1}{2}\alpha(\phi_n H^m{}_l+\phi_l H^m{}_n),\\&\Delta^5_{mn}=\tfrac{1}{2}\alpha\left(\frac{\partial\phi_n}{\partial x^m}+\frac{\partial\phi_m}{\partial x^n}\right)-\alpha\phi_p\Gamma^p_{mn}+\tfrac{1}{2}\alpha^3\phi_p\phi_m H^p{}_n+\tfrac{1}{2}\alpha^3\phi_p\phi_n H^p{}_m.\end{aligned}\right\}. \quad (11)$$

The expressions $(H^p{}_m)$ are mixed tensor components derived from the field intensity components

$$H_{mn}=\frac{\partial\phi_m}{\partial x^n}-\frac{\partial\phi_n}{\partial x^m}.$$

The analogy that has been drawn with the electromagnetic theory gives the form of the field equations (2). These consist of a set

$$T^{\mu\nu}{}_{;\nu}=0, \quad \ldots\ldots\ldots\ldots \quad (12)$$

where summation is implied over ν from 1 to 5 and there are five equations corresponding to $\mu=1\ldots5$. The symbol $;\nu$ denotes covariant differentiation.

It will be assumed that the tensor, T, is in the form of a curl of a vector. Thus the covariant components are of the form

$$T_{\mu\nu}=\frac{\partial\psi_\nu}{\partial x_\mu}-\frac{\partial\psi_\mu}{\partial x^\nu}. \quad \ldots\ldots\ldots \quad (13)$$

It will be assumed that only four components (ψ_μ) exist, the reason for this limitation being that a similar function occurs in the quantum theory and that four components have been found sufficient. Advantage is therefore taken of this simplification.

The coordinate x^5 will be supposed to occur in the form already assumed, but in order to take account of the possibility of the necessity of including a mass similar to μ_0 as described in Part I., it will be assumed that the operation $\frac{\partial}{\partial x^5}=\frac{2\pi\iota\mu_0 c}{h}$ and for convenience $\frac{2\pi\iota\mu_0 c}{h}$ will be denoted by σ. As before, α will be used to denote e/m_0c^2.

Thus $$T_{55}=0 \quad \text{since} \quad \psi_5=0, \qquad (14)$$

and $$T_{m5}=-T_{5m}=-\frac{\partial\psi_m}{\partial x^5}=-\sigma\psi_m. \qquad (15)$$

Similar relations hold for the conjugate quantities and, remembering that $\sigma^*=-\sigma$,

$$T^*_{m5}=\sigma\psi^*_m, \qquad (16)$$

ψ can be treated as a vector in four or five dimensions owing to the absence of ψ_5.

Thus $$\psi^m=\gamma^{m\nu}\psi_\nu=\gamma^{mn}\psi_n=g^{mn}\psi_n. \qquad (17)$$

Equation (12) when written in full is

$$\frac{\partial T^{\mu\nu}}{\partial x^\nu}+\Delta^\nu_{\alpha\nu}T^{\alpha\nu}+\Delta^\nu_{\alpha\mu}T^{\mu\alpha}=0,$$

i. e., $$\frac{1}{\sqrt{\gamma}}\frac{\partial}{\partial x^\nu}(\sqrt{\gamma}T^{\mu\nu})=0, \qquad (18)$$

since the second term vanishes on account of the antisymmetry of T. With the values of the components $(\gamma_{\mu\nu})$ introduced the determinant, γ, composed of these components has the same value as the corresponding quantity in the theory of relativity and is accordingly placed equal to unity.

Thus in the case $\mu=m$, (18) becomes

$$\frac{\partial T^{mn}}{dx^n}+\sigma T^{m5}=0. \qquad (19)$$

With the same notation as before,

$$T^{mn}=A^{mn},$$

$$T^{m5}=A^m{}_0-\alpha\phi_n A^{mn}.$$

But $$T_{m5}=A_{m0} \quad (cf.\ (9)).$$

Thus $$A_{m0}=-\sigma\psi_m \quad \text{and} \quad A^m{}_0=-\sigma\psi^m.$$

Equation (19) therefore becomes

$$\left(\frac{\partial}{\partial x^n}-\alpha\sigma\phi_n\right)A^{mn}=\sigma^2\psi^m. \qquad (20)$$

A^{mn} can be expressed in terms of the components of ψ.

By (9)

$$T_{mn}=\frac{\partial\psi_n}{\partial x^m}-\frac{\partial\psi_m}{\partial x^n}=A_{mn}+\alpha\sigma(\phi_m\psi_n-\phi_n\psi_m).$$

Thus

$$A^{mn}=g^{mr}g^{ns}A_{rs}=g^{mn}g^{ns}\left(\frac{\partial\psi_s}{\partial x^r}-\frac{\partial\psi_r}{\partial x^s}\right)-\alpha\sigma(\phi^m\psi^n-\phi^n\psi^m)$$

$$=g^{mr}g^{ns}\left(\frac{\partial}{\partial x^r}-\alpha\sigma\phi_r\right)\psi_s-g^{mr}g^{ns}\left(\frac{\partial}{\partial x^s}-\alpha\sigma\phi_s\right)\psi_r \quad . \quad . \quad . \quad (21)$$

In the same way the equation for the conjugate tensor $T^{*\mu\nu}$ is

$$T^{*\mu\nu}{}_{;\nu}=0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

and leads to

$$\left(\frac{\partial}{\partial x^n}+\alpha\sigma\phi_n\right)A^{*mn}=\sigma^2\psi^{*m}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

where

$$A^{*mn}=g^{mr}g^{ns}\left(\frac{\partial}{\partial x^r}+\alpha\sigma\phi_r\right)\psi^*_s-g^{mr}g^{ns}\left(\frac{\partial}{\partial x^s}+\alpha\sigma\phi_s\right)\psi^*_r. \quad . \quad . \quad (24)$$

In applications to the problems of the quantum theory gravitational effects are negligible and the components, (g^{mr}), are zero or unity according as $m\neq r$ or $m=r$.

It remains to consider the case $\mu=5$ in equation (12). In this case

$$T^{5\nu}{}_{;\nu}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

and since $T^{55}=0$, this reduces to

$$\frac{\partial T^{5n}}{\partial x^n}=0.$$

It follows that

$$\frac{\partial}{\partial x^n}(A_0{}^n-\alpha\phi_p A^{pn})=0,$$

and therefore

$$\sigma\frac{\partial\psi^n}{\partial x^n}-\alpha\frac{\partial\phi_p}{\partial x^n}A^{pn}-\alpha\phi_p\frac{\partial A^{pn}}{\partial x^n}=0.$$

By means of equation (20) this reduces to

$$\left(\frac{\partial}{\partial x^n}-\alpha\sigma\phi_n\right)\psi^n=-\frac{1}{2}\frac{\alpha}{\sigma}H_{mn}A^{mn}. \quad . \quad . \quad . \quad . \quad . \quad (26)$$

Similarly, for the conjugate quantities

$$\left(\frac{\partial}{\partial x^n}+\alpha\sigma\phi_n\right)\psi^{*n}=\frac{1}{2}\frac{\alpha}{\sigma}H_{mn}A^{*mn}. \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Equations (20), (23), (26), and (27), which are four dimensional forms of (12) and (22), are the field equations of this theory of the electric charge.

2. *The Equation satisfied by the Components of ψ.*

In the case of the electromagnetic equations the vector and scalar potentials are introduced in the same way as the vector ψ. These potential components satisfy a well-known wave equation, and by analogy it follows that the ψ-components satisfy a similar equation. To obtain this equation (18) may be written in the form

$$\frac{\partial}{\partial x^{\nu}}(\gamma^{\mu\alpha}\gamma^{\nu\beta}\mathrm{T}_{\alpha\beta})=\gamma^{\mu\alpha}\gamma^{\nu\beta}\left(\frac{\partial^2\psi_\beta}{\partial x^\nu\partial x^\alpha}-\frac{\partial^2\psi_\alpha}{\partial x^\nu\partial x^\beta}\right)+\left(\frac{\partial\psi_\beta}{\partial x^\alpha}-\frac{\partial\psi_\alpha}{\partial x^\beta}\right)\frac{\partial}{\partial x^\nu}(\gamma^{\mu\alpha}\gamma^{\nu\beta})=0. \quad . \quad . \quad . \quad (28)$$

It is unnecessary to proceed with any further reduction of this equation in its general form since its character can be appreciated from the important case of the absence of a gravitational field.

In this case γ^{rs} is zero when $r\neq s$ and is unity otherwise.

Let the case, $\mu=m$, be examined.

The second term on the right-hand side is

$$-\gamma^{\mu\alpha}\gamma^{\nu\beta}\frac{\partial^2\psi_\alpha}{\partial x\ \partial x^\beta}=-\sum_n\frac{\partial^2\psi_m}{\partial x^{n2}}+2\alpha\sigma\phi^n\frac{\partial\psi_m}{\partial x^n}-\sigma^2(1+\alpha^2\phi_r\phi^r)\psi_m.$$

This expression is familiar in Schroedinger's theory of the charged particle. It will be written in the form $-\Box\psi_m$, for it is minus the Laplacian of ψ_m in five dimensions.

The remaining terms may be treated similarly, the first term giving

$$\left(\frac{\partial}{\partial x^m}-\alpha\sigma\phi^m\right)\sum_n\frac{\partial\psi_n}{\partial x^n}-\alpha\sigma\phi^n\frac{\partial\psi_n}{\partial x^m}+\alpha^2\sigma^2\phi^m\phi^n\psi_n,$$

and the last term may be simplified to become $-\alpha\sigma\psi_n\dfrac{\partial\phi^m}{\partial x^n}$.

These two sets of terms together give

$$\left(\frac{\partial}{\partial x^m}-\alpha\sigma\phi_m\right)\sum_n\left(\frac{\partial}{\partial x^n}-\alpha\sigma\phi_n\right)\psi_n+\alpha\sigma\mathrm{H}_{mn}\psi_n,$$

there being no longer any need to distinguish between covariant and contravariant components since $\gamma^{rs}=\delta_{rs}$, where $\delta_{rs}=0$ or 1 according as $r\neq s$ or $r=s$.

Thus by means of equation (26), (28) reduces to

$$\Box\psi_m=\alpha\sigma\mathrm{H}_{mn}\psi_n-\frac{1}{2}\frac{\alpha}{\sigma}\left(\frac{\partial}{\partial x^m}-\alpha\sigma\phi_m\right)\mathrm{H}_{pn}\mathrm{A}_{pn}. \quad . \quad . \quad . \quad (29)$$

When no electromagnetic field is present this equation reduces to

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\frac{\partial^2}{\partial z^2}-\frac{1}{c^2}\frac{\partial^2}{\partial t^2}\right)\psi_m=\frac{4\pi^2\mu_0^2c^2}{h^2}\psi_m. \quad . \quad . \quad . \quad (30)$$

This is the well-known equation occurring in the wave mechanical theory of the free electron of mass μ_0.

A similar equation holds for ψ_m^*.

3. *The Energy Tensor.*

This tensor is formed by proceeding further with the analogy with the electromagnetic theory. It will be seen that just as the electric charge is included in a generalized momentum vector so the current density is part of the generalized energy tensor, which thus becomes the energy, momentum, and current tensor.

In order to include the tensor $(T^{\mu\nu})$ and its conjugate $(\overset{*}{T}{}^{\mu\nu})$ symmetrically in the structure of the field tensor a new tensor

$$P^{\mu}{}_{\nu}=T^{*\mu\alpha}T_{\nu\alpha}+T^{\mu\alpha}T^{*}_{\nu\alpha} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

is adopted.

The coordinate x^5 does not occur in $P^{\mu}{}_{\nu}$ on account of its structure as the sum of the products of conjugate quantities.

The field tensor, $R^{\mu\nu}$, is now formed according to the relation

$$R^{\mu\nu}=P^{\mu\nu}-\tfrac{1}{4}\gamma^{\mu\nu}P, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where

$$P=P^{\mu}{}_{\mu}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

and is described as the energy tensor.

A law is postulated according to which

$$R^{\mu\nu}{}_{;\nu}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

A convenient form of this equation for the present purpose is

$$R_{\nu}{}^{\mu}{}_{;\nu}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

for in the case $\mu=5$

$$R_{5}{}^{\nu}{}_{;\nu}=\frac{\partial R_{5}{}^{n}}{\partial x^{n}}+\Delta^{\alpha}_{5\nu}R_{\alpha}{}^{\nu}=0,$$

taking account of the fact that $R_5{}^{\nu}$ does not contain x^5 and that $\gamma=1$.

By making use of the values of the coefficients $\Delta^{\alpha}_{5\nu}$ given by (11) and of the fact that $R^{\nu\sigma}$ is symmetric, it follows that $\Delta^{\alpha}_{5\nu}R_{\alpha}{}^{\nu}$ vanishes.

Thus

$$\frac{\partial R_{5}{}^{n}}{\partial x^{n}}=0.$$

This represents the conservation of the vector with four components, $R_5{}^n$, $(n=1\,..\,4)$ and suggests the identification of this quantity with the current vector, which will be assumed to be equal to $R_5{}^n$ multiplied by a constant.

From (32) $\qquad R_{5}{}^{n}=\gamma_{5\alpha}R^{\alpha n}=\gamma_{5\alpha}P^{\alpha n}=P_{5}{}^{n},$

and since $T_{55}=0$, $\qquad R_{5}{}^{n}=T^{*nr}T_{5r}+T^{nr}T^{*}_{5r}.$

Expressing the right-hand side in terms of four dimensional quantities

$$\begin{aligned} R_{5}{}^{n}&=A^{*nr}A_{0r}+A^{nr}A^{*}_{0r}\\ &=\sigma(\psi_{r}A^{*nr}-\psi^{*}_{r}A^{nr}). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)\end{aligned}$$

Thus the current vector is proportional to

$$X^n = \psi_r A^{*nr} - \psi_r^* A^{nr}. \qquad (37)$$

Another component of the energy tensor which may be easily calculated is R^{mn}.

From equation (34)

$$\begin{aligned} R^{m\nu}{}_{;\nu} &= \frac{\partial R^{m\nu}}{\partial x^\nu} + \varDelta^m_{\alpha r} R^{\alpha\nu} \\ &= \frac{\partial R^{mn}}{\partial x^n} + \varGamma^m_{an} R^{an} + (\varDelta^m_{an} - \varGamma^m_{an}) R^{an} + 2\varDelta^m_{5n} R^{5n} \\ &= R^{mn}{}_{;n} - \alpha H^m{}_n R_5{}^n, \end{aligned}$$

where $R^{mn}{}_{;n}$ denotes four dimensional covariant differentiation of R^{mn}, which can be regarded as a four-tensor by (5).

Equation (34) is thus equivalent to

$$R^{mn}{}_{;n} = \alpha H^m{}_n R_5{}^n = \alpha\sigma H^m{}_n X^n \qquad (38)$$

for the case $\mu = m$.

The right-hand side is equal to a constant multiplied by the Lorentz force. Thus R^{mn} is proportional to the energy-momentum tensor of the field. This justifies the adoption of $R^{\mu\nu}$ for the purpose of generalization of this tensor, all components of which can be interpreted. It combines in one expression the energy, momentum and current all of which are expressible in terms of the four components (ψ^m) and the conjugate (ψ^*_m).

It is found that

$$P^{mn} = A^{*mr} A^n{}_r + A^{mr} A^{*n}{}_r - \sigma^2(\psi^{*m}\psi^n + \psi^m\psi^{*n}), \qquad (39)$$

$$P = A^{*mn} A_{mn} + A^{mn} A^*_{mn} - 4\sigma^2 \psi^{*m}\psi^m, \qquad (40)$$

and

$$R^{mn} = P^{mn} - \tfrac{1}{4} g^{mn} P. \qquad (41)$$

The energy tensor is now defined quantitatively by the introduction of the constant factor $h^2c^2/4\pi^2$ and is denoted by E with components

$$E^{mn} = \frac{h^2c^2}{4\pi^2} R^{mn}. \qquad (42)$$

Neglecting the gravitational field the expression for the energy density is

$$\begin{aligned} E_{44} = \frac{h^2c^2}{4\pi^2} (A^*_{23}A_{23} + A^*_{31}A_{31} + A^*_{12}A_{12} + A^*_{14}A_{14} + A^*_{24}A_{24} + A^*_{34}A_{34}) \\ + \mu_0^2 c^4 (\psi_1^*\psi_1 + \psi_2^*\psi_2 + \psi_3^*\psi_3 + \psi_4^*\psi_4). \qquad (43) \end{aligned}$$

This energy consists of a part dependent upon the electromagnetic field and of a part independent of it. The expression thus expresses the fact that there is an intrinsic energy associated with the charge. It is to be noted that this quantity is essentially positive; there are no negative energy values.

The quantitative expression for the current vector can be found from equation (38), for in terms of E^{mn} it must take the form

$$E^{mn}{}_{;n}=H^{m}{}_{n}J^{n},$$

where J^n is a component of the current vector J.

Hence by inserting the values of the constants

$$\begin{aligned} J^n &= \iota\frac{hec}{2\pi}\frac{\mu_0}{m_0}(\psi_r A^{*nr}-\psi_r^* A^{nr}), \\ &= \frac{h^2c^2}{4\pi^2}\alpha R_5{}^{n}. \qquad (44) \end{aligned}$$

Like the energy tensor this expression has a contribution independent of the external electromagnetic field which may be obtained by writing the components $(\phi_s)=0$. This contribution may be regarded as arising from an intrinsic electromagnetic property of the charge. Its magnitude is

$$\iota\frac{hec}{2\pi}\frac{\mu_0}{m_0}\left\{\psi_r^2\frac{\partial}{\partial x^n}\left(\frac{\psi_r^*}{\psi_r}\right)+\frac{\partial}{\partial x^r}(\psi_r^*\psi_n-\psi_r\psi_n^*)\right\}. \qquad (45)$$

The latter term may be described as arising from an electromagnetic-moment tensor, M, and an electromagnetic-moment of magnitude M may be ascribed to the charge, a component of M being

$$M_{rn}=\iota\frac{hec}{2\pi}\frac{\mu_0}{m_0}(\psi_r^*\psi_n-\psi_r\psi_n^*).$$

It is to be noted that although a change in sign of the current vector may occur, the energy density, as mentioned above, is always positive. In this theory of the electric charge not only is there no difficulty about the sign of the energy but the correct value results without reference to electron structure and without difficulties concerning infinite self energy.

The results obtained should be compared with those deduced by Proca from a different point of view and with which they are in agreement. The remarks on energy values and electromagnetic moment have already been made by him ⁽³⁾.

This theory is not a quantum theory of the charge or of the field. The only appeal made so far to the quantum theory is in the use of the Wilson-Sommerfeld conditions when an application was made of them in order to incorporate the quantization of electric charge into the theory. The next step is to proceed to quantization, and this process may be regarded as the introduction of matrices under the guidance of the mathematical form which has been developed. The quantization of the field theory has already been made by analogy with the procedure adopted in the electromagnetic theory by the adoption of Proca's expression for the energy value ⁽⁴⁾.

A beginning will now be made with a process which may be described as a quantization of geometrical concepts.

The geometrical theory already discussed may be regarded as the basis of the theory applicable in macroscopic physics, and a limitation to ordinary space-time has been suggested by it.

In order to appeal to geometrical ideas in the description of microscopic phenomena a new conception of length will be introduced in the form of a matrix length.

4. *Quantization and Matrix Geometry.*

The well-known quantization of motion is expressed by means of the Wilson-Sommerfeld conditions of the older form of the quantum theory. These conditions become

$$pq-qp=h/2\pi i$$

in the new form of the theory, where p is the momentum matrix and q the matrix of the conjugate of p.

In the case of the momentum component π_5, and the coordinate, x^5, it is to be expected that the relation will become

$$\pi_5 x^5-x^5\pi_5=h/2\pi i, \qquad (46)$$

where the symbols now denote matrices.

Instead of π_5 the foregoing theory suggests the substitution of the matrix N with $\pi_5=\mathrm{N}l/\alpha c$.

The matrix N represents the number of charges present and has proper values 0, 1, 2, etc.

Generalized dynamics suggests the replacement of x_5 by a matrix ω to correspond to a variable with unit periodicity.

Thus the substitution $x^5=\dfrac{h}{m_0 c}\omega$ is made.

Consequently equation (46) becomes

$$\mathrm{N}\omega-\omega\mathrm{N}=\frac{1}{2\pi i}. \qquad (47)$$

From this it may be deduced that

$$\mathrm{N}e^{2\pi i\omega}-e^{2\pi i\omega}\mathrm{N}=e^{2\pi i\omega}, \qquad (48)$$

$$\mathrm{N}e^{-2\pi i\omega}-e^{+2\pi i\omega}\mathrm{N}=-e^{-2\pi i\omega}. \qquad (49)$$

These equations are already well known in the quantum theory, and it will be observed that

$$e^{2\pi i\omega}\mathrm{N}=(\mathrm{N}-1)e^{2\pi i\omega},$$

and more generally for certain functions $f(\mathrm{N})$

$$e^{2\pi i\omega}f(\mathrm{N})=f(\mathrm{N}-1)e^{2\pi i\omega}. \qquad (50)$$

The operator $e^{2\pi\iota\omega}$ may be regarded as one which changes the number of particles present in a particular form, and it has been used in the description of certain phenomena where transitions take place between fundamental particles.

Similarly,

$$e^{-2\pi\iota\omega}f(\mathrm{N})=f(\mathrm{N}+1)e^{-2\pi\iota\omega}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{51}$$

The Matrix Length (5).

The foregoing example shows in a simple case how matrices are introduced into the theory of physical phenomena which take place within dimensions far below those considered in macroscopic physics. Conceptions with regard to measurement of intervals of length and time may well require modification in these conditions, and a limitation with regard to localization of a particle in space-time has already been suggested.

Guided by such suggestions it is proposed to introduce the element of length into the theory by means of a matrix instead of by means of the expressions for $d\sigma^2$ or ds^2.

In ordinary analysis the element of length is expressed in the form of a vector by

$$ds=\mathbf{i}\,dx+\mathbf{j}\,dy+\mathbf{k}\,dz,$$

where (**ijk**) are the three fundamental unit vectors. It will be supposed that in matrix geometry these vectors are replaced by matrices. The present purpose is served by the introduction of five matrices of which γ_μ is typical. This matrix is supposed to possess a covariant character and to be associated with a contravariant companion related to it in the usual way $\gamma^\mu=\gamma^{\mu\nu}\gamma_\nu$. Thus the matrix is capable of operations as a matrix and as a vector.

A matrix length $d\sigma$ is defined by means of

$$d\sigma=\gamma_\mu dx^\mu, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{52}$$

where summation over Greek suffixes like μ is again from 1 to 5. On squaring this matrix.

$$d\sigma^2=\gamma_\mu\,dx^\mu\gamma_\nu\,dx^\nu=\Sigma\gamma_\mu^2\,(dx^\mu)^2+\Sigma(\gamma_\mu\gamma_\nu+\gamma_\nu\gamma_\mu)\,dx^\mu\,dx^\nu,$$

where, in the second summation, μ and ν are unlike.

By the assumptions

$$\left.\begin{aligned}\gamma_\mu^2&=\gamma_{\mu\mu},\\ \gamma_\mu\gamma_\nu+\gamma_\nu\gamma_\mu&=2\gamma_{\mu\nu},\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \tag{53}$$

the squared element becomes equal to that already introduced. Thus the square of the matrix length is equal to the square of the ordinary length. The relations (53) are described as Tetrode's relations since he was the first to introduce them, his application being to four-dimensional geometry. In the quantum theory, as in the cases already pointed out, a number of quantities appear which are not identified in their character of five-dimensional quantities. One of the most important of these is the matrix α_0 of the quantum equation. It is similar in character to A_0, which was introduced earlier in this paper and which is a scalar quantity appearing as a substitute for the covariant vector component C_5. Thus as in equations (2) and (3)

$$\gamma_5=\alpha_0, \quad \ldots\ldots\ldots\ldots \quad (54)$$

and for the contravariant component

$$\gamma^5=\alpha_0-\alpha\phi_m\gamma^m, \quad \ldots\ldots\ldots\ldots \quad (55)$$

α denoting the constant already introduced.

From the value of γ_{55} it follows that

$$(\alpha_0)^2=(\gamma_5)^2=\gamma_{55}=1. \quad \ldots\ldots\ldots\ldots \quad (56)$$

The contravariant matrices (γ^m) can be placed equal to components (α^m) of a four-dimensional character as in the case of vector components. The following relations hold :

$$\left.\begin{aligned}(\gamma^\mu)^2&=\gamma^\mu\gamma^\mu,\\ \gamma^\mu\gamma^\nu+\gamma^\nu\gamma^\mu&=2\gamma^{\mu\nu},\\ \gamma^\mu\gamma_\nu+\gamma_\nu\gamma^\mu&=2\delta^\mu_\nu,\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (57)$$

where δ^μ_ν is zero unless $\mu=\nu$ and is then unity.

A special case of the last relation is

$$\gamma^m\gamma_5+\gamma_5\gamma^m=0,$$

since $m\neq5$, and this can be put into the form

$$\alpha^m\alpha_0+\alpha_0\alpha^m=0,$$

when it is recognized as an important relation between the matrices which occur in the quantum equation.

Other relations between these equations can be obtained by (57).

Thus

$$\left.\begin{aligned}\alpha_m\alpha_n+\alpha_n\alpha_m&=2g_{mn},\\ \alpha^m\alpha^n+\alpha^n\alpha^m&=2g^{mn}.\end{aligned}\right\} \quad \ldots\ldots\ldots\ldots \quad (58)$$

The matrix length can now be written in the form

$$d\sigma=\gamma_\mu dx^\nu=\alpha_m dx^m+\alpha_0 d\theta. \quad \ldots\ldots\ldots\ldots \quad (59)$$

In the same way any vector may be associated with a matrix according to the relation

$$C=\gamma_\mu C^\mu=\alpha_m A^m+\alpha_0 A_0. \quad \ldots\ldots\ldots\ldots \quad (60)$$

The principle now assumed is that in the appeal to geometrical conceptions for a description of microscopic physics the matrix length and matrix vectors take the place of the ordinary lengths and vectors employed in macroscopic physics.

5. *Parallel Displacement in Matrix Geometry* ⁽⁶⁾.

The change of components of a vector (C^μ) when it undergoes a parallel displacement is defined according to the following rule:

$$dC^\mu = -\Delta^\mu_{\lambda\nu} C^\lambda dx^\nu. \quad \text{(61)}$$

In terms of the covariant components

$$dC_\mu = \Delta^\lambda_{\mu\nu} C_\lambda dx^\nu. \quad \text{(62)}$$

The coefficients ($\Delta^\mu_{\lambda\nu}$) are, as before, the Christoffel bracket expressions for the five-dimensional continuum. In consequence there is no change of length in the vector C as a result of a parallel displacement in this continuum.

The question of the change in the matrix length, $\gamma^\mu C_\mu$ ($=\gamma_\mu C^\mu$), when the vector, C_μ, undergoes a parallel displacement will now be considered.

A matrix, ψ, will be supposed to exist such that the quantity

$$L = \gamma^\mu C_\mu \psi \quad \text{(63)}$$

undergoes no charge when the vector is subject to a parallel displacement.

This matrix, ψ, will be regarded as a matrix of four components ($\psi_{11}, \psi_{21}, \psi_{31}, \psi_{41}$) which may be denoted by ($\psi_1, \psi_2, \psi_3, \psi_4$). It is, moreover, scalar in character, having none of the vector properties of the γ-matrices.

The change in γ^μ resulting from the displacement will be denoted by $\frac{\partial \gamma^\mu}{\partial x^\nu} dx^\nu$, and thus the condition imposed upon L leads to

$$\frac{\partial \gamma^\mu}{\partial x^\nu} dx^\nu C_\mu \psi + \gamma^\mu \Delta^\lambda_{\mu\nu} C_\lambda dx^\nu \psi + \gamma^\mu C_\mu \frac{\partial \psi}{\partial x^\nu} dx^\nu = 0.$$

This is to be true for all vectors and for all displacements.

Thus
$$\left(\frac{\partial \gamma^\mu}{\partial x^\nu} + \gamma^\rho \Delta^\mu_{\rho\nu}\right)\psi + \gamma^\mu \frac{\partial \psi}{\partial x^\nu} = 0. \quad \text{(64)}$$

Write
$$\frac{\partial \gamma^\mu}{\partial x^\nu} + \gamma^\rho \Delta^\mu_{\rho\nu} = K^\mu{}_\nu. \quad \text{(65)}$$

It will be assumed that a law affecting $K^\mu{}_\nu$ exists which may be described as the quantum law.

A very simple law would be $K^\mu{}_\nu = 0$, but this relation is too stringent for our purpose. For in order to satisfy it, it would be necessary to make ψ a constant which would rule out the idea of introducing ψ as a gauging function for the continuum.

The position is similar to that which arose in the case of the choice of a law of gravitation. In this case the law concerned a limitation on the Riemann-Christoffel tensor, but the simplest law, which was the vanishing of this tensor, would have cut away the basis of the geometrical explanation of the theory of gravitation. Instead of this the law of gravitation was obtained by means of a limitation upon the contracted tensor.

If the suggestion from this analogy be followed, the quantum law or law of gauging which is to be adopted is

$$K^{\mu}{}_{\mu}=0. \qquad \qquad (66)$$

This, by equation (64), leads to the following equation for ψ:

$$\gamma^{\mu}\frac{\partial\psi}{\partial x^{\mu}}=0. \qquad \qquad (67)$$

It is supposed that x^5 occurs in the various functions used in the same simple way as before. In this case equation (67) for the case of the electron becomes

$$\gamma^{m}\frac{\partial\psi}{\partial x^{m}}+\frac{2\pi i m_0 c}{h}\gamma^5\psi=0.$$

When this is expressed in terms of the matrices α^m and α_0, it becomes

$$a^{m}\left(\frac{h}{2\pi i}\frac{\partial}{\partial x^{m}}-\frac{e}{c}\phi_m\right)\psi+m_0 c\alpha_0\psi=0, \qquad \qquad (68)$$

which is the well-known form of the quantum equation of the electron. It is possible to put $K^{\mu}{}_{\nu}$ into a particular form by means of the relation

$$\gamma^{\mu}\gamma^{\nu}+\gamma^{\nu}\gamma^{\mu}=2\gamma^{\mu\nu}$$

and of the identity

$$\frac{\partial\gamma^{\mu\nu}}{\partial x^{\sigma}}+\Delta^{\mu}_{\alpha\sigma}\gamma^{\alpha\nu}+\Delta^{\nu}_{\alpha\sigma}\gamma^{\mu\alpha}=0.$$

Combining this with (65)

$$(K^{\mu}{}_{\sigma}\gamma^{\nu}+\gamma^{\nu}K^{\mu}{}_{\sigma})+(K^{\nu}{}_{\sigma}\gamma^{\mu}+\gamma^{\mu}K^{\nu}{}_{\sigma})=0.$$

This relation is satisfied identically if $K^{\mu}{}_{\sigma}$ is of the form

$$K^{\mu}{}_{\sigma}=\Delta_{\sigma}\gamma^{\mu}-\gamma^{\mu}\Delta_{\sigma}, \qquad \qquad (69)$$

where Δ_{σ} is a matrix.

It will be noted that the γ-matrices vary according to (65). This may be regarded as a definition of parallel displacement for these matrices.

If C'_{μ} denote a component of the vector after parallel displacement

$$C'_{\mu}=C_{\mu}+\Delta^{\lambda}_{\mu\nu}C_{\lambda}dx^{\nu},$$

and the vector with these components at the adjacent point may be said to be equal to the original vector (C_{μ}) at the initial point. Applying this idea to the matrices, the matrix (γ'^{μ}) at the neighbouring point equivalent to (γ^{μ}) initially is given by

$$\gamma'^{\mu}=\gamma^{\mu}+(K^{\mu}{}_{\nu}-\gamma^{\rho}\Delta^{\mu}_{\rho\nu})\,dx^{\nu}.$$

In this case the change in value of the matrix length $\gamma^{\mu}C_{\mu}$, which will be denoted by l is $K^{\mu}{}_{\nu}C_{\mu}dx^{\nu}$.

According to the value of $K^{\mu}{}_{\nu}$ given in equation (69) this change may be expressed in the form

$$dl=(\Delta_{\nu}l-l\Delta_{\nu})dx^{\nu}. \qquad (70)$$

6. *Projective Matrix Geometry.*

In the matrix introduced by Einstein containing a definition of distant parallelism a group of quantities (h_{ma}) is introduced, sixteen in number, with m and n able to assume the integral values 1 to 4. These are point functions and they satisfy the relations

$$\left.\begin{aligned} h_{ma}h_{na}&=g_{mn},\\ h^m_a h^n_a&=g^{mn},\\ h_{ma}h^n_a&=\delta^n_m,\\ h^m_a h_{mb}&=\delta_{ab}.\end{aligned}\right\} \qquad (71)$$

It will be noted that m and n denote the variant character of h while a and b are in the character of labels.

It may be assumed that at each point of space a local system of unit orthogonal vectors (ϵ_a) exists. There is also a set of curvilinear vectors (i^m) which are chosen at any point so that the scalar product $(i^m . i^n)$ is equal to g^{mn}. This takes the place of the choice of perpendicular reference vectors in ordinary vector analysis.

Any vector A can be expressed in terms of its covariant components (A_m) in the form (7)

$$A=A_m i^m. \qquad (72)$$

The notation this gives for the square of the vector A,

$$A^2=g^{mn}A_mA_n.$$

It is necessary for the purpose of the representation of a vector in terms of contravariant components to represent the curvilinear vectors alternatively by vectors (i_m) chosen so that the scalar product $(i_m . i_n)$ is equal to g^{mn}. Thus the vector A may be represented by

$$A=A^m i_m. \qquad (73)$$

In addition to the scalar product already introduced there occurs

$$i_m i^n=\delta^n.$$

For the present purpose this notation will be employed in a five-dimensional continuum. The foregoing relations are taken over by using γ for g and the affixes μ and ν for m and n.

It is possible to carry out operations applied to the neighbourhood of a point of the curvilinear space either by reference to the curvilinear vectors or to the local straight vectors (ϵ_a). These satisfy the relations

$$\epsilon_a . \epsilon_b=\delta_{ab}. \qquad (74)$$

A vector A may be represented either as having curvilinear components (A^μ or A_μ) or straight components (A_a).

The quantities (h^μ_α) are used to relate the two types of component and they appear in the rôle of projection coefficients. The relations are

$$A_\alpha = h_{\mu\alpha} A^\mu = h^\mu_\alpha A_\mu . \qquad (75)$$

$$A^\mu = h^\mu_\alpha A_\alpha . \qquad (76)$$

$$A_\mu = h_{\mu\alpha} A_\alpha . \qquad (77)$$

The square of the length, A^2, is expressible in two ways which must be equal.

$$A^2 = \gamma_{\mu\nu} A^\mu A^\nu = \gamma_{\mu\nu} h^\mu_\alpha h^\nu_\beta A_\alpha A_\beta .$$

But $$A^2 = \Sigma A_\alpha^2 = \delta_{\alpha\beta} A_\alpha A_\beta ,$$

from which it is concluded that

$$\gamma_{\mu\nu} h^\mu_\alpha h^\nu_\beta = h^\mu_\alpha h_{\mu\beta} = \delta_{\alpha\beta} .$$

This agrees with the relations (71).

Again, from (75),

$$\Sigma A_\alpha^2 = h^\mu_\alpha h^\nu_\alpha A_\mu A_\nu ,$$

whence $$h^\mu_\alpha h^\nu_\alpha = \gamma^{\mu\nu} ,$$

$$h_{\mu\alpha} h_{\nu\alpha} = \gamma_{\mu\nu} .$$

The vector (i^μ) has the single component $A_\mu = 1$, the others are zero. Thus $A_\alpha = h^\mu_\alpha$ and the corresponding "flat" vector has components ($h^\mu_1 \ldots h^\mu_5$) or the vector corresponding to i^μ is $h^\mu_\alpha \epsilon_\alpha$. Similarly, the components for the vector ϵ_α are zero except for $A_\alpha = 1$. The corresponding contravariant components are (h^μ_α) or the equivalent of ϵ_α is $h^\mu_\alpha i_\mu$ or similarly $h_{\mu\alpha} i^\mu$.

The line element in curvilinear space has components (dx^μ) and, if the corresponding element in flat space be (dz_α), it follows that

$$\left.\begin{aligned} dz_\alpha &= h_{\mu\alpha}\, dx^\mu , \\ dx^\mu &= h^\mu_\alpha\, dz_\alpha . \end{aligned}\right\} \qquad (78)$$

If the scalar product of two curvilinear vectors be considered it is easy to prove that

$$A_\mu B^\mu = A_\alpha B_\alpha .$$

These ideas may be taken over into matrix geometry by the use of local orthogonal matrices (E_α) related to the curvilinear matrices (γ^μ) by similar relations.

Curvilinear quantities, *e. g.*, matrix lengths and matrix vectors, represented by $\gamma_\mu dx^\mu$ and $\gamma_\mu A^\mu$, can be represented locally by corresponding equivalent quantities in flat space, $E_\alpha dz_\alpha$ and $E_\alpha A_\alpha$, the relations (75) to (77) holding for the vector components.

The matrix γ^μ is represented by $h^\mu_\alpha E_\alpha$ in flat space, and conversely E_α is represented by $h_{\mu\alpha}\gamma^\mu$ or $h^\mu_\alpha \gamma_\mu$ in curved space. The correspondence of two quantities in the two representations will be denoted by the symbol →. Thus $\gamma^{\mu} \rightarrow h^{\mu}_{\alpha} \mathrm{E}_{\alpha}$.

The relation
$$\gamma^{\mu}\gamma^{\nu}+\gamma^{\nu}\gamma^{\mu}=2\gamma^{\mu\nu},$$
in terms of the E-matrices becomes
$$h^{\mu}_{\alpha}h^{\nu}_{\beta}(\mathrm{E}_{\alpha}\mathrm{E}_{\beta}+\mathrm{E}_{\beta}\mathrm{E}_{\alpha})=2\gamma^{\mu\nu}.$$
and on account of the properties of the h-coefficients it follows that
$$\mathrm{E}_{\alpha}\mathrm{E}_{\beta}+\mathrm{E}_{\beta}\mathrm{E}_{\alpha}=2\delta_{\alpha\beta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (79)$$

In this case, as in the case of vector representation, it is impossible to make a general representation of curvilinear quantities in flat space. The representation is possible locally in a space "tangential" to the curved space.

It is important to adopt a convention with regard to scalar quantities such as the matrix ψ.

In the representation a point (x^{μ}) of curved space is associated with a point (z_a) of flat space, and the dual representation is made in the small region about (x^{μ}) and (z_a). If a scalar function ψ exists at (x^{μ}) the corresponding function in the flat space will be taken to be the same function of the z-coordinates as the first function is of the x-coordinates.

Thus
$$\psi(x^{\mu}) \rightarrow \psi(z_{\alpha}),$$
and for the neighbouring point
$$\psi(x^{\mu}+dx^{\mu}) \rightarrow \psi(z_{\alpha}+dz_{\alpha}).$$
Thus
$$\frac{\partial\psi}{\partial x^{\mu}}\,dx^{\mu} \rightarrow \frac{\partial\psi}{\partial z_{\alpha}}\,dz_{\alpha}.$$

This is an example of the equivalence that exists in the case of
$$\mathrm{A}^{\mu}\mathrm{B}_{\mu} \rightarrow \mathrm{A}_{\alpha}\mathrm{B}_{\alpha}.$$

The product $\gamma^{\mu}\dfrac{\partial\psi}{\partial x^{\mu}}$ in the same way becomes $\mathrm{E}_{\alpha}\dfrac{\partial\psi}{\partial z_{\alpha}}$, and thus there is a local representation of the quantum equation of the same form as the original curvilinear form.

These points have been recently elaborated and discussed in a paper by H. W. Haskey (8).

References.

(1) Rosenfeld, *Acad. Roy. Belgique*, xiii. p. 447 (1927).
(2) Fisher and Flint, Proc. Roy. Soc. A, cxxvi. p. 645 (1929).
(3) Proca, *J. Phys. Radium*, vii. p. 347 (1936); and viii. p. 23 (1937).
(4) Fröhlich, Heitler, and Kemmer, Proc. Roy. Soc. A, clvxi. p. 154 (1938).
(5) Flint, Proc. Roy. Soc. A, cl. p. 441 (1935). Mimura, J. Sci. Hiroshima Univ. A, v. p. 99 (1935).
(6) Flint, Proc. Phys. Soc. l. p. 899 (1938).
(7) Flint, Proc. Roy. Soc. A, ciii. p. 644 (1923).
(8) Haskey. In the press.

XLIV. *A Rigorous Theory of the Recombination of both small and large Ions in Gases at High Pressures.*

By W. R. Harper, Ph.D., Research Associate, H. H. Wills Physical Laboratory, University of Bristol *.

[Received November 17, 1939.]

Introduction.

No completely rigorous treatment of any recombination problem has so far been presented. The purpose of this paper is to deal with the limited case of the later stages of the recombination of ions at high pressures, taking into account all the physical factors involved, and calculating the recombination coefficient α by a method that is logically precise. The relevant physical processes being well understood, it will be possible to develop a theory analogous to the kinetic theory of gases, which can be used with similar confidence for the interpretation of experimental results.

Hitherto, in default of a generally accepted theory, it has been necessary to choose between the predictions of different theories by reference to experiment. This has recently been done for low pressures by Sayers (1938), and for high pressures by Mächler (1936). The results for high pressures were in conformity with the measurements of Lea (1934), and with earlier results cited in a previous paper of the author's (1932), though representing a great improvement in accuracy. But for comparison with theory the accuracy of the measurement is unfortunately not very significant because no determinations of mobility were made on the same specimens of gas, and it is really the ratio of recombination coefficient to mobility that is predicted by theory. Nevertheless, the work of Mächler has shown that experiment decides in favour of the formula deduced by Langevin (1903), and quite probably with an accuracy of a few per cent.

Now this is very difficult to understand, because elementary considerations lead one to the conclusion that Langevin's theory must give a value for α that is considerably too low. It has been pointed out by Townsend (1915) that Langevin's theory is incomplete in that it takes no account of the effect of diffusion in bringing ions of opposite sign together. It only allows for the effect of electrostatic attraction. It happens, as a result, that no mention has ever to be made of the size of the ions in calcu-

---

* Communicated by the Author.

lating α, which depends only on the mobilities. It follows that, for given mobilities, a large ion is endowed with the same recombination coefficient as a small ion, and this is clearly incorrect. We may be sure that diffusion contributes to α for a large ion, and the following argument suggests that it must also do so for a small ion. It is generally accepted that a pair of ions may be considered as being effectively recombined when they have approached to within the distance at which their energy of thermal agitation is equal to their mutual negative electrostatic energy, and it may easily be calculated (*e. g.*, from the results of the present paper) that the number of encounters of this kind that would arise from diffusion alone, is comparable with the number that would arise from electrostatic attraction alone. Since, in fact, both electrostatic attraction and diffusion are operative, it would appear to follow that the value of α as calculated by Langevin from the effect of electrostatic attraction alone must be considerably too low.

The explanation, which we shall obtain from a rigorous treatment, is somewhat remarkable. It will appear that diffusion is about seven times less effective when combined with electrostatic attraction, as when acting alone. It furthermore happens, only because the law of attraction is that of the inverse square, that the resulting small contribution from diffusion is exactly balanced by our elementary argument having neglected the small chance of the ions separating after having supposedly become effectively recombined. For this reason, our general formula for any size of ion will reduce to exactly that of Langevin for the case of small ions.

Enough should already have been said to make it apparent that the mechanism of recombination phenomena is a much more subtle affair than has generally been supposed. The author has previously had occasion (1936 *b*) to refer to the necessity for mathematical rigour in obtaining unambiguous results in the analogous problem of coagulation. The rigorous treatment of ordinary recombination will lead to some conclusions on the mechanism of the initial stages of recombination, showing that certain current ideas derived from physical intuition are the precise opposite of the truth.

An aspect of recombination which has received undeservedly little attention is the question of the validity of the law that the rate of recombination is equal to αn^2, where n is the concentration of ions of either sign—in other words, whether or not the value of α is independent of the concentration. It was recently stated that the law has been proved theoretically and experimentally verified. Actually, neither is the case. Sutherland (1909) has pointed out that early theories have done little more than beg the question, and that more careful considerations lead one to expect some other law. The author, in an earlier treatment (1932) of recombination at high pressures, in which all the physical factors were taken into account, but in which the mathematical methods depended on plausibility rather than on proof, derived a formula * for α which had a slight dependence on n, α being greater for smaller concentrations. A slight variation of this kind would be masked in an experiment if the ions were not homogeneous with respect to recombination coefficient. The ions of smallest recombination coefficient would be left until last, thus compensating for the increase in α due to decreased concentration. It is always difficult to be sure exactly what ions are being experimented upon, and it is probable that no experiments have yet been performed in which the possibility here contemplated has definitely been ruled out. It is rather the contrary. Both Mächler and Sayers used air, and the latter concluded on experimental grounds that his ions were not homogeneous, and quoted other experimental evidence in favour of this view. We shall therefore investigate under what conditions the αn^2 law is valid. We shall be able, not merely to justify its use in cases where it is normally applied, but also to show that it is applicable to conditions where at first sight it would seem absurd to apply it.

General Considerations.

In developing a theory of recombination we have to choose between two mathematical lines of attack. In the one (*e. g.*, the approach of J. J. Thomson) attention is focused on the way in which a typical ion moves about as a result of thermal agitation and electrostatic forces, and the rapidity with which it searches out other ions, and makes encounters of a kind which will lead to recombination with them, calculated. The physical details of the mechanism are thus clearly correlated with the mathematical formulation. In the other line of attack (*e. g.*, in the theory of Langevin) attention is concentrated on the spatial *distribution* of the ions, and the way in which it alters with time. The correlation with physical detail is less apparent, and indeed, the results obtained depend on factors which, from the first standpoint, at first seem irrelevant. In using the second method, therefore, it is essential to justify with some care its application to the actual physical problem. Failure to do so in the face of an apparent incompatibility with the first standpoint was the basis of certain criticisms of the method advanced in previous papers of the author (1932, 1934, 1936 *a*), but it has appeared since (Fuchs, 1936, and Harper, 1936 *b*) how these criticisms can be met. We shall employ the second line of attack in this paper, following the treatment of Townsend, but it will be justified and interpreted in terms of the first.

---

* The formula also had what then seemed the virtue of giving numerical values about half as large again as those of Langevin's.

No ionizing agent produces a random distribution of ions. For this reason it is customary to allow time for adequate mixing to be brought about by diffusion before determining α. We are concerned primarily with this late stage in the recombination process, which we shall call ordinary recombination, but the rate at which it occurs will prove to be, in one sense, a legacy from what has preceded it. We must therefore consider, to some extent, the initial stages. Before the ions have become well mixed there will be localized regions where the concentration is higher than the average, and other regions where it is lower, resulting on the whole in an abnormally high recombination. It is sometimes permissible to treat this state of affairs as a case of ordinary recombination by considering the regions of different concentration separately. But it has been recognized for some time that there is yet another kind of recombination which the author (1933) has termed " preferential recombination." It may happen, particularly at high pressures, that pairs of ions are formed with their positive and negative members very close together. These will then be attracted to each other with electrostatic forces so great as to ensure there being only a small chance of their diffusing apart and recombining with other ions. They have, in fact, a marked preference for recombining with each other. Therefore, it has been said, this form of recombination must be completed before ordinary recombination can become established. We shall confirm the existence of a form of recombination which does lead to the rapid disappearance of very close pairs before ordinary recombination sets in, and shall continue to call it preferential recombination, but we shall find it must be explained in quite other terms. For the present, however, the explanation does not concern us. It is only necessary to allow for the possibility of the existence of preferential recombination from the beginning of a rigorous treatment of ordinary recombination because its existence is ultimately predicted and explained by that treatment.

The reason why preferential recombination is of importance in dealing with ordinary recombination is that we shall calculate α from a knowledge of the precise spatial distribution of the ions in ordinary recombination, and this distribution is a legacy from the preferential recombination that preceded it, and is not, as might easily have been supposed, a strictly random one. In a random distribution the probability of finding an ion in a small element of volume at a given distance from any chosen ion is independent of the distance, and the same for all the ions. In the actual distribution the probability is zero for distances less than the critical radius for capture, and may be reduced for still greater distances. This will be the case if, in setting up the distribution we have to consider, preferential recombination has led to the disappearance of ions with neighbours closer than the average, leaving an excess of ions with neighbours further away than the average, so that the ions which remain are surrounded, on the average, by a partially denuded region. The volume effectively denuded round each ion in this way is quite small, so that, for sufficiently small ionic concentrations, the total volume denuded in this way is still small compared with the total volume of the gas. The actual distribution will then only differ from a random one in that a small number of ions which would be present in a random one are absent in it. We shall calculate α for this case, subsequently investigating the conditions for the assumption to be valid. We shall find that it is only when the difference between the actual distribution and a random one is negligible in the above sense that the αn^2 law holds, but, on the other hand, it is only a knowledge of the departure from a random distribution which will enable α to be calculated.

Not having a theory of preferential recombination, we do not know, in advance, what precisely the departure from a random distribution will be. But the required distribution is merely that to which *any* initial distribution tends asymptotically, as a result of diffusion, electrostatic attraction and recombination. We may therefore take any convenient initial distribution for which we can solve the subsequent motion, ignore what happens at the start, and consider only its behaviour after the lapse of that time during which preferential recombination has changed the initial distribution into the relevant one. It will be convenient to take as initial distribution a strictly random one.

The Effect of Concentration on α.

We shall defer certain difficulties connected with mutual attraction by considering for a moment the analogous problem of coagulation. This only differs in that now the particles move about under the influence of diffusion alone, coagulating after approach to within a definite critical distance. We might suppose that in this case there would be nothing corresponding to preferential recombination, but this supposition would later prove to be erroneous, so we must continue to bear in mind its effects. Compare now such a physical system with a fictitious one consisting of a random distribution of mathematical points subject to the same law of motion as the particles, and each surrounded by a hypothetical sphere whose radius can later be identified with the critical distance for capture. The point centres will move along Brownian tracks which at times will intersect the spheres without being affected thereby. Owing to the tortuous nature of the tracks these intersections will be multiple. We shall count a multiple intersection as a single collision since, in the corresponding physical system, we envisage a single encounter as sufficient to lead to coagulation. The collision frequency for the interpenetrating system does not change with time, there being no coagulation, and for this reason is easier to calculate than the rate of coagulation in the physical system. Consider a typical sphere. In unit time there is a certain probability that any given centre will have visited the sphere, and the collision frequency is just this probability summed for all the centres and multiplied by the number of the spheres. It is clearly therefore proportional to the square of the concentration of the centres. In considering any centre as a candidate for visiting the typical sphere, we contemplated it as capable of passing without hindrance through any other sphere it might encounter on the way. Had such an encounter led to coagulation, we should have had to multiply each individual probability of visiting the typical sphere by the probability that the centre in question got through without coagulation. Now this probability is nearly zero for nearly all the centres, because there are so many spheres with which a given centre may coagulate, and in general no reason for it to prefer our chosen one. The complication is associated with the fact that the concentration is changing, and with the possibility that the collision frequency is proportional to some function, other than the square, of the concentration. This possibility will be a reality unless, for a time during which the concentration has changed by only a small fraction, the only centres contributing appreciably to the collision frequency are specially favoured ones which are very unlikely to encounter another sphere. Otherwise the collision frequency will be multiplied by a factor depending on the concentration and representing the average probability that a centre gets through to the chosen sphere.

We shall however now show that, provided the number of centres which would disappear in changing a random distribution to the proper one is small compared with the number which disappear during a small change in concentration of the system, we may identify the collision frequency in the interpenetrating system with the instantaneous coagulation rate in the physical system. That we are entitled to do so may be seen most easily from a consideration of what would happen if our mathematical model were suddenly given physical reality. There would immediately be some preferential coagulation but, as we have postulated, a negligible amount. The tracks of the remainder of the particles would be unaffected until they suffered collision. After a lapse of time sufficient for a large number but a small fraction of the particles to have coagulated, the number that would have coagulated would be less than the number that would have made interpenetrating collisions for two reasons. Firstly, if we follow a centre in the interpenetrating system and it makes a collision, it may not do so (as a particle) in the other system, owing to the previous disappearance of the supposedly encountered particle. But the chance of this occurring is less than the fraction of the particles which has already disappeared by coagulation, and this we have postulated as small. Secondly, some of the collisions counted in the interpenetrating system are collisions of a centre which has already collided with another sphere, and these should not be counted. But again our postulate implies that only a small fraction can be of this kind. The collision frequency calculated for the interpenetrating system is therefore applicable to the physical system during a small change in concentration of the latter, and this conclusion holds for any stage of the coagulation. We are therefore justified in applying the calculated collision frequency to the coagulation process by merely making the concentration a function of the time.

If now we turn to the problem of recombination in which mutual electrostatic attraction modifies the effect of thermal agitation, we see that even for ions which have no size, that is for an interpenetrating system, there is recombination. This follows from the fact that the inverse square law provides each centre with an infinitely deep potential hole into which a centre of the opposite sign can be permanently captured, since kinetic energy, in excess of the thermal value, will be lost at high pressures by collisions with gas molecules. The author's approximate theory must now be rejected on the grounds that it envisaged a stationary state of an interpenetrating system, which we see to be impossible.

Consider therefore an infinite space of volume V, containing n_1 ions per c.c., of mobility k_1 and charge $+e$, distributed strictly at random through it, and n_2 ions per c.c., of mobility k_2 and charge $-e$, similarly distributed. For the present we shall think of the ions as points subject to thermal agitation, and attracting (or repelling) each other according to the inverse square law. We have to calculate the rate at which ions will recombine during a small change in concentration of the system. The actual distribution of ions about any positive ion is discontinuous, but if we construct another system whose distribution about an origin is the sum of the actual distributions about each positive ion, then the distribution of centres in this system is uniform and continuous. If furthermore the centres have the same motion as they had relative to the appropriate positive ion in the actual distribution, then the rate at which the negative centres arrive at the origin for the first time is the rate at which ions disappear in the first system. This follows since, if any negative centre arrives at the origin in the second system, then one negative ion will have arrived at some positive ion in the first system, and that positive ion will almost certainly still be available for recombination, since we have postulated a small change only in concentration before it occurs.

The concentration of negative centres in the second system is $n_1 n_2 \mathrm{V}$. They move along Brownian tracks characterized (Harper, 1936 *b*) by a diffusion coefficient D_1+D_2, and modified by the attractions and repulsions of the *first* system. A given ion is influenced by

(*a*) the attraction of the ion at the origin,

(*b*) other ions which it encounters during the time considered,

(*c*) ions which remain distant during this time.

The resultant effect of (*c*) is a small fluctuating electrostatic field which will alter the diffusion coefficient, but by a negligible amount. The effect of (*b*) is twofold. In the first place the encounter may lead to recombination. We have seen, however, that the effect of this is duly allowed for by making the concentration a function of the time. In the second place the encounter may lead to a complete change in path. If now the disturbing ion is surrounded by a uniform distribution at large distances, its influence will change the distribution in its immediate neighbourhood, but will leave the distribution at large distances unaffected. It follows that although the change of path is of importance when considering individual ions, the *distribution* is only affected in the neighbourhood of the disturbing ion, or when the disturbing ion is near the origin where the concentration is not uniform because of preferential recombination. The rate of arrival of centres at the origin is therefore unaffected by (*b*) if the probability is small that a disturbing ion be near the origin. The sufficient (and it may easily be seen necessary) condition for this to be so is that the particulate volume be large compared with the volume within which the ionic distribution has been appreciably disturbed by preferential recombination. It is included in the similar condition already postulated. When this condition is fulfilled, we are left with (*a*), we may treat the individual centres of the second system as mutually independent and only affected by the field of the origin, and α will not depend on the concentration. Our conclusion is that α is independent of the concentration unless this is sufficiently high for preferential recombination to have been appreciable. The exact significance of this restriction will be discussed later.

The Exact Theory.

We have now simplified the problem to a problem in diffusion. We have to consider a uniform and continuous distribution of independent centres distributed throughout infinite space with a concentration n_1n_2V, and each attracted to the origin with a force proportional to the inverse square of the distance. The centres diffuse about with a diffusion coefficient D_1+D_2, and we require to know the rate at which centres visit the origin for the first time. This is equal to the rate at which recombinations occur in the volume V. The correct procedure for obtaining the result

(Harper, 1936 *b*) is to solve the appropriate diffusion equation for the boundary condition of concentration equal to zero in the neighbourhood of the origin, and calculate the rate at which centres are absorbed there. We may include the case of small ions recombining with large ions by putting the concentration equal to zero for $r=a$, where a is the radius of the large ion. The equation to be solved is

$$\frac{dn}{dt}=\frac{1}{r^2}\frac{\partial}{\partial r}\left\{(\mathrm{D}_1+\mathrm{D}_2)r^2\frac{\partial n}{\partial r}+\frac{(k_1+k_2)en}{\mathrm{K}}\right\},$$

with $n=n_1n_2\mathrm{V}$ everywhere for $t=0$, and $n=0$ at $r=a$ for $t>0$. K is the dielectric constant, inserted so as to include the case of a liquid.

Unfortunately a general analytical solution of this equation cannot be obtained by ordinary methods. It is, however, apparent that initially there is a rapid absorption of centres at $r=a$, which then falls off asymptotically to a steady rate. The initial absorption corresponds to preferential recombination, and the final rate to ordinary recombination. We therefore obtain the ordinary recombination coefficient from the equation

$$(\mathrm{D}_1+\mathrm{D}_2)\frac{\partial}{\partial r}\left\{r^2\frac{\partial n}{\partial r}\right\}+\frac{(k_1+k_2)e}{\mathrm{K}}\frac{\partial n}{\partial r}=0.$$

The solution of this is

$$n=\frac{n_1n_2\mathrm{V}\left[1-\exp\left\{\frac{(k_1+k_2)e}{(\mathrm{D}_1+\mathrm{D}_2)\mathrm{K}}\left(\frac{1}{r}-\frac{1}{a}\right)\right\}\right]}{1-\exp\left\{-\frac{(k_1+k_2)e}{(\mathrm{D}_1+\mathrm{D}_2)\mathrm{K}a}\right\}}.$$

The rate of arrival of centres at $r=a$ is

$$4\pi a^2(\mathrm{D}_1+\mathrm{D}_2)\left.\frac{\partial n}{\partial r}\right|_{r=a}=\frac{4\pi n_1n_2\mathrm{V}e(k_1+k_2)}{\mathrm{K}\left[1-\exp\left\{-\frac{(k_1+k_2)e}{(\mathrm{D}_1+\mathrm{D}_2)\mathrm{K}a}\right\}\right]},$$

whence

$$\alpha=\frac{4\pi e(k_1+k_2)}{\mathrm{K}\left[1-\exp\left\{-\frac{(k_1+k_2)e}{(\mathrm{D}_1+\mathrm{D}_2)\mathrm{K}a}\right\}\right]}.$$

We shall use the well-known proportionality between mobility and diffusion coefficient by writing $(k_1+k_2)e/3\mathrm{K}(\mathrm{D}_1+\mathrm{D}_2)=\sigma_0$, giving finally

$$\alpha=\frac{4\pi e(k_1+k_2)}{\mathrm{K}\left\{1-\exp\left(-\frac{3\sigma_0}{a}\right)\right\}},$$

in which $\sigma_0=2{\cdot}03\,.\,10^{-6}\times 273/\mathrm{K}\theta$ cm. for a temperature of $\theta°$ abs. It is to be remembered that e and k must be expressed in the same units.

The author (1932) has shown that σ_0 is the value of r at which the radial velocity of escape by diffusion is equal to the inward drift in the field for the average ion, and has therefore used it as the critical radius for capture for small ions. It is easy to show that it is also the distance at which the mutual electrostatic energy of the two ions is equal to their total energy of thermal agitation, *i. e.*, that $e^2/K\sigma_0=3B\theta$, where B is Boltzmann's constant. We do not, however, require the conception of a critical radius for capture in applying the present theory to small ions, because, as the size of the ion is reduced, the formula for α tends rapidly to Langevin's result of $4\pi e(k_1+k_2)/K$, and has practically attained this value for a radius much larger than that of a small ion.

In the absence of electrostatic attraction the formula for α reduces to $4\pi a(D_1+D_2)$, and in the absence of diffusion to $4\pi e(k_1+k_2)/K$. The former is Smoluchowski's coagulation coefficient, discussed in previous papers of the author (1934, 1936 *a* and *b*), and the latter is, of course, Langevin's recombination coefficient.

The Diffusion Paradox.

We are now in a position to reconsider our criticism of Langevin's neglect of diffusion. We argued that when the ions had approached to within a distance σ_0 they were effectively recombined. If this were accurately true α would be given by the general formula with $a=\sigma_0$ instead of with $a=0$. The ratio of these two values of α therefore gives the probability of recombination after approach to within the distance σ_0. It is 0·95, so our supposition was not far wrong. We then argued that diffusion alone would lead to a number of σ_0 collisions comparable with the number due to electrostatic attraction alone. Reference to the formulæ shows that it would lead to one-third as many. We concluded that when both were operative the collision frequency would be markedly higher than with electrostatic attraction alone, and hence that Langevin's formula must be too low. We now see that our conclusion was wrong because of the exponential term in the general formula. This results in the collision frequency being only slightly increased by diffusion in the presence of electrostatic attraction even though separately it would have a much larger effect. Furthermore this slight increase is exactly balanced by the small chance of escape from a σ_0 collision. This surprising deduction from a rigorous theory could hardly have been anticipated from approximate reasoning.

Limits of Validity.

Experiment shows that the formula breaks down for pressures below about 10 atmospheres. There are three reasons why the theory is no longer applicable at low pressures :—

(*a*) As previously pointed out by the author (1932), when the mean free path is comparable with σ_0, diffusion theory is no longer competent to give the rate of arrival of ions at the sphere of radius σ_0 (after which capture is almost certain at high pressures).

(*b*) Ions can separate after approaching to within the distance σ_0 owing to the retention of excess kinetic energy. Such excess energy is not admitted by the equation used for calculating α.

(*c*) The mobility may no longer be taken as constant in the region where it matters, *i. e.*, outside the sphere σ_0. At 10 atmospheres the ratio of the field at σ_0 in volts per cm. to the pressure in mm. of Hg is about 5, and it is for such values that Mitchell and Ridler (1934) find the breakdown occurring.

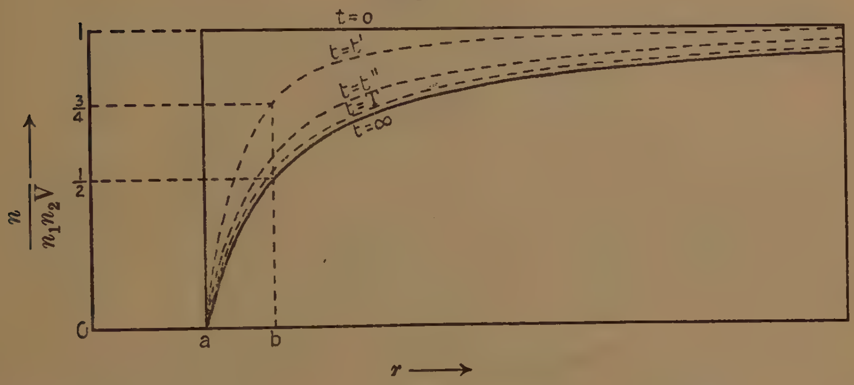
We have seen that the theory also breaks down when the concentration is so high that the number of ions which would disappear in changing a random distribution to the actual one is comparable with the number which disappear during a small change in concentration of the system. Preferential recombination is then no longer limited to the initial stage of the recombination process, but continues to affect the rate of recombination until the concentration has fallen below this limit. Now preferential recombination is the abnormally high recombination that is due to ions of opposite sign being very close together. As already mentioned, it has hitherto been supposed to result from the intense electrostatic attraction between the ions, diffusion keeping the effect within finite limits. Actually, as we can now show, the precise opposite is the case. It is an effect which *results* from diffusion, and is *non-existent* when electrostatic attraction alone is operative. Its magnitude is determined by the number of centres that are absorbed before a stationary state is established. The initially high rate of absorption arises from the *diffusion* term, the discontinuity in concentration making this infinite for $t=0$. In the absence of diffusion we have Langevin's solution which is a uniform concentration both initially and finally, and there is no preferential recombination because there is no initial unstationary state. To pursue further the implications of these conclusions is outside the scope of this paper, but we shall estimate the magnitude of preferential recombination when both diffusion and electrostatic attraction have to be considered, in order to find at what concentrations the theory breaks down. It will appear that the theory is valid for remarkably high concentrations.

Having no analytical solution of the equation, we must supplement such useful mathematical information as we possess by what is obvious from physical considerations. In fig. 1 the known initial and asymptotic distributions have been plotted (for $a=3\sigma_0$) as full line curves. Since the one turns into the other as a result of diffusion and drift, it is clear that all intermediate distributions must be convex upwards throughout, the distribution for t_2 lying entirely within the distribution for t_1. They must therefore have the general form of the dotted curves.

The number of centres which are absorbed at $r=a$ before the asymptotic distribution is attained is infinite, so we shall estimate the number which are absorbed before the distribution has changed to within a small fraction γ of its asymptotic value. Call this distribution the final state, and let it be attained after time T. If we denote by J the rate at which centres cross the sphere of radius r at time t we have

$$J=4\pi\left\{(D_1+D_2)r^2\frac{\partial n}{\partial r}+\frac{ne}{K}(k_1+k_2)\right\},$$

Fig. 1.



and at time T, J will have approached to within the order of γ of its asymptotic value. It follows furthermore from the fact that the curves are convex upwards that

$$\frac{\partial n}{\partial r}<\frac{n_1n_2V}{r-a},$$

so J is always and everywhere less than J′ where

$$J'=4\pi n_1n_2V(D_1+D_2)\left\{\frac{r^2}{r-a}+3\sigma_0\right\}.$$

J itself is infinite for $t=0$ and $r=a$. To avoid considerations of convergence therefore, we estimate the number of centres that cross a sphere of radius b greater than a in setting up the final state, and say that the number S absorbed at $r=a$ is less than this number together with the number that was contained between the two spheres at $t=0$. It will be convenient to choose b as the value of r for which the asymptotic curve has risen to half value, when we have

$$\frac{1}{b}=\frac{1}{a}+\frac{1}{3\sigma_0}\log\frac{1}{2}\left\{1+\exp\left(-\frac{3\sigma_0}{a}\right)\right\},$$

and b is always of the order of magnitude of a, irrespective of the magnitude of σ_0.

Now

$$S<\tfrac{4}{3}\pi n_1 n_2 V(b^3-a^3)+J_6'T,$$

and t is related to n by the equation

$$t_2-t_1=4\pi b^2\int_{t_1}^{t_2}\frac{dn_b}{\left.\frac{\partial J}{\partial r}\right|_b},$$

Fig. 2.

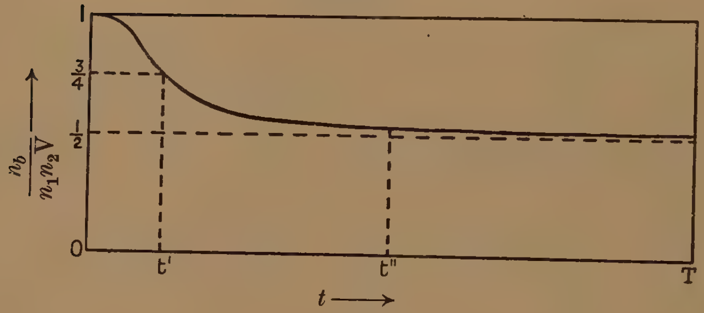
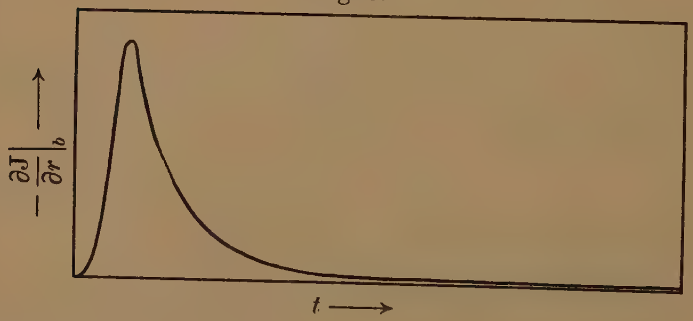


Fig. 3.



which comes from writing our original equation in the form

$$\frac{dn}{dt}=\frac{1}{4\pi r^2}\frac{\partial J}{\partial r}.$$

It is clear from physical considerations that n_b varies with time in the general manner shown in fig. 2, whence $\left.\frac{\partial J}{\partial r}\right|_b$ has the general form shown in fig. 3. Once again we must avoid convergence difficulties at $t=0$.

We define t' as the value of t for which n_b has fallen halfway to its asymptotic value, and let $\frac{1}{2}\mathrm{T}=t''$. It is clear from fig. 2 that $t'<t''$. This allows us to write

$$\mathrm{T}=8\pi b^2\int_{t''}^{\mathrm{T}}\frac{dn_b}{\left.\frac{\partial \mathrm{J}}{\partial r}\right|_b}<8\pi b^2\int_{t'}^{\mathrm{T}}\frac{dn_b}{\left.\frac{\partial \mathrm{J}}{\partial r}\right|_b}.$$

We may bring $\left.\frac{\partial \mathrm{J}}{\partial r}\right|_b$ outside the integral if we substitute its minimum absolute value, which occurs at the upper limit, when n_b has fallen to within γ of its asymptotic value. We do not know how n_b approaches its asymptote, and this determines the required value of $\left.\frac{\partial \mathrm{J}}{\partial r}\right|_b$. However, we may put the latter equal to γ^p times its maximum absolute value, and be sure that p is quite a small power. Consider now

$$\left.\frac{\partial \mathrm{J}}{\partial r}\right|_b=4\pi(\mathrm{D}_1+\mathrm{D}_2)\left\{b^2\left.\frac{\partial^2 n}{\partial r^2}\right|_b+(2b+3\sigma_0)\left.\frac{\partial n}{\partial r}\right|_b\right\}.$$

The first term is negative, and the second positive, the sum being negative or zero. The two terms are equal and cancel each other for the asymptotic curve, and are unequal but of the same order of magnitude for the intermediate curves, as is apparent from fig. 1. $\left.\frac{\partial n}{\partial r}\right|_b$ is of the order of $\frac{n_1n_2\mathrm{V}}{2(b-a)}$ for the intermediate curves, so the maximum value of $\left.\frac{\partial \mathrm{J}}{\partial r}\right|_b$ is of the order of

$$-\frac{2\pi n_1n_2\mathrm{V}(\mathrm{D}_1+\mathrm{D}_2)(2b+3\sigma_0)}{b-a}.$$

We therefore have

$$\begin{aligned}\mathrm{T}&<\frac{4b^2(b-a)}{\gamma^p n_1n_2\mathrm{V}(\mathrm{D}_1+\mathrm{D}_2)(2b+3\sigma_0)}\int_{\mathrm{T}}^{t'}dn_b\\&<\frac{4b^2(b-a)}{\gamma^p n_1n_2\mathrm{V}(\mathrm{D}_1+\mathrm{D}_2)(2b+3\sigma_0)}\int_{\infty}^{t'}dn_b\\&<\frac{b^2(b-a)}{\gamma^p(\mathrm{D}_1+\mathrm{D}_2)(2b+3\sigma_0)}.\end{aligned}$$

It follows that

$$\mathrm{S}<\tfrac{4}{3}\pi n_1n_2\mathrm{V}(b^3-a^3)+\frac{4\pi n_1n_2\mathrm{V}b^2\{b^2+3\sigma_0(b-a)\}}{\gamma^p(2b+3\sigma_0)},$$

which is always of the order of $4\pi n_1n_2\mathrm{V}a^3/\gamma^p$, whatever the value of σ_0. The theory is valid if S is small compared with a small fraction of $n_1\mathrm{V}$ or $n_2\mathrm{V}$, that is if $\mathrm{S}<\gamma^2\mathrm{V}\sqrt{n_1n_2}$, say, or, finally, if $4\pi a^3\sqrt{n_1n_2}<\gamma^{p+2}$. Inspection of the mathematics suggests that we may be a little more precise, and say that $4\pi a^3\sqrt{n_1 n_2}=\gamma^{p+2}$ is, very approximately, the condition that the theory is breaking down to the extent of giving a small fractional error of γ in α. Remembering that p is probably about 1 or 2, we may express our conclusion most simply as follows :—

The theory breaks down when the diameter of an ion is no longer small compared with the average separation of the ions, and the anticipated fractional error in α is of the order of magnitude of the ratio of these distances.

The diameter of a small ion is of atomic dimensions, so we deduce that the theory is valid for small ions and high pressures up to extremely high concentrations. It does not break down on this account even for the enormous concentration of ions in an α particle track at 100 atmospheres. The recombination coefficient would however be somewhat increased in such an extreme case by the mobility breakdown assuming a new importance.

Summary.

The recombination coefficient for both large and small ions at high pressures is derived by a rigorous treatment. For small ions the formula reduces to that of Langevin. It is shown how it was possible for Langevin to obtain the correct result without allowing for the effects of diffusion. The condition for the recombination coefficient to be independent of concentration is investigated, and it is found that for small ions the theory is not invalidated on this account unless concentrations are greatly in excess of those found in an α particle track at 100 atmospheres.

References.

Fuchs, N., Trans. Faraday Soc. xxxii. p. 1146 (1936).

Harper, W. R., Proc. Camb. Phil. Soc. xxviii. p. 219 (1932) (see also *ibid.* xxxi. p. 429 (1935) for a correction); Proc. Camb. Phil. Soc. xxix. p. 149 (1933); Trans. Faraday Soc. xxx. p. 636 (1934) (see also *ibid.* xxxi. p. 774 (1935) for a correction); Trans. Faraday Soc. xxxii. p. 1139 (1936 *a*); *ibid.* xxxii. p. 1148 (1936 *b*).

Langevin, P., *Ann. de Phys.* xxviii. p. 433 (1903).

Lea, D. E., Proc. Camb. Phil. Soc. xxx. p. 80 (1934).

Mächler, W., *Z. Phys.* civ. p. 1 (1936).

Mitchell, J. H., and Ridler, K. E. W., Proc. Roy. Soc. A, cxlvi. p. 911 (1934).

Sayers, J., Proc. Roy. Soc. A, clxix. p. 83 (1938).

Sutherland, W., Phil. Mag. xviii. p. 341 (1909).

Townsend, J. S., ‘Electricity in Gases’ (Oxford University Press), chapter vi. (1915). *Cf.* also Fuchs, N., *Z. Phys.* lxxxix. p. 736 (1934).

XLV. *The Fundamental Principles of Energetics* *.

By J. N. Brønsted (Institute of Physical Chemistry, the University of Copenhagen) †.

[Received October 10, 1939.]

	Page
1. Introduction	449
2. The Basic Process	450
The Concept of Work and the Work Principle	450
3. Entropy and Temperature	453
4. The Concept of Heat and Energetic Heat Evolution	454
5. Energy	458
6. Heat and Work in Traditional Thermodynamics	459
7. The Fundamental Principles of Energetics	465
8. Summary	468

1. *Introduction.*

The science of thermodynamics in its usual traditional and classical formulation originates historically in the important investigations which were carried out about the middle of the last century on the relations between the concepts of heat and work. The chief problem was to combine the theory of the "motive power of heat," which had been brilliantly developed by Carnot, with the principle of the equivalence of heat and work which had been advanced by Mayer, Colding, and Joule. The nature of the problem to be solved in this field was clearly formulated by William Thomson. It is to Clausius that credit is usually given for a definite and final solution of these problems.

There can, in fact, be no doubt that the theory of Clausius provides a satisfactory basis for carrying out thermodynamic calculations and for the quantitative correlation of energetic phenomena. Closer analysis does, however, undoubtedly show that there are decisive weaknesses in the fundamental ideas of this theory with regard to the genetic relation between heat and work, and that even its interpretation of the concept of heat evolution and absorption cannot be logically maintained. It may be said in general that the weaknesses referred to are due to the fact that emphasis has been laid on the principle of the equivalence of heat and work in certain simple processes at the cost of the fundamental idea of Carnot. This has led to the suppression of important aspects of the subject which could otherwise have been developed as a fundamental generalization of Carnot's principle.

---

* The present paper is a short summary of two papers written in Danish: J. N. Brønsted, "Om Relationen mellem Varme og Arbejde," *Det. Kgl. Danske Vid. Selsk. Medd.* xv. no. 4 (1937), and "De thermodynamiske Hovedsætningers Grundlag og Formulering," *ibid.* xvi. no. 10 (1939). The translation has been made by R. P. Bell.

† Communicated by the Author.

It is the object of the present paper to attempt to elucidate the relation between heat and work on a broader and more rational basis than has so far been the case. As already mentioned, this problem is closely connected with a more careful analysis of the concept of heat. Moreover, it will be necessary to extend the concept of work beyond the rather narrow limits implied in classical thermodynamics, and in general to give a more precise formulation to the fundamental concepts and phenomena associated with energy changes. For this reason we shall consider our problem as dealing with " energetics " in its general sense rather than with the more circumscribed field of " thermodynamics."

2. *The Basic Process.*

The Concept of Work and the Work Principle.

The problem of the rational formulation of the fundamental laws of energetics can be approached in a number of different ways, each of which is equally legitimate. We shall begin by pointing out that a number of simple natural processes can be described as the transport of a quantity from one state to another, where in using the word " transport " we are pre-supposing that the amount of the entity transported undergoes no change. For example, this will be true of the movement of a weight perpendicular to the earth's surface, since the mass of the weight (which in this case represents the quantity) is regarded as remaining unchanged. It will also be true for the transference of a volume from one state to another, since the volume which is received by the one state is taken as being equal to the volume lost by the other state. It should also be noticed that in these processes we can discriminate between the two directions, since in the absence of external forces the weight will move downwards, and under the same conditions the volume will move from a lower to a higher pressure. We shall term the direction of the spontaneous process positive, and that of the reverse process negative.

We shall next point out as a matter of experience that simple processes of this kind can be coupled together to give reversible compound processes in which the two quantities concerned are transported in opposite directions. In describing the transport as reversible, we mean that any state whatever which is conditioned by the coupling can be reached without the occurrence of any changes other than the coupled transport processes. We can also say that the process takes place through equilibrium states.

We shall now generalize the above considerations by stating that all macroscopic natural phenomena are conditioned by the transport of *energetic quantities* between different states, and that in the reversible case the changes consist entirely of transports of this kind. Such a transport of an energetic quantity will be termed a *basic energetic process.* If a process as a whole is reversible, this demands (as in the special cases already considered) a coupling between the basic processes involved which regulates their concurrent progress, one taking place in a positive and the other in a negative direction. As before, the sign of the direction is determined by observing the transport after the coupling has been removed, the basic processes in that case being positive.

As already mentioned, mass and volume are important energetic quantities, but in addition to these we have entropy, quantity of a chemical substance, quantity of electricity, and physical surfaces and lengths. It is not possible to give any general *proof* of the transportability of these quantities, but the reversibility principle permits us to use this assumption without contradiction. In other words, this principle enables us to define a quantity as transportable without contradiction. This important point will be considered further in section (3) in connexion with the definition of entropy.

The concept of work is very intimately connected with fundamental energetic processes. It is possible to associate with any fundamental process an *amount of work,* which is by definition proportional to the amount of the quantity transported and also depends on the two states between which the transport takes place. The work for a basic process between given initial and final states can thus be defined by the equation

$$\delta A = (P_1 - P_2)\delta K, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where δK is the quantity transported and P_1 and P_2 magnitudes which depend only on the initial and final states respectively. These magnitudes we shall call *potentials.* Quantities and potentials are mutually conjugate, each pair corresponding to a particular kind of work.

Equation (1) defines the amount of work associated with a fundamental process between given initial and final states. If two such processes are coupled together reversibly, then, if $\delta A'$ and $\delta A''$ are the corresponding infinitesimal work terms, it is found experimentally that the relation

$$\delta A' + \delta A'' = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

is always valid, independent of the path by which the corresponding quantities pass between the initial and final states of the basic processes. This means that changing the path in one of the processes will not alter the quantity that can be reversibly transported. The numerical values of $\delta A'$ and $\delta A''$ will of course depend upon the units employed for the quantities and the potentials. Hence, in order that equation (2) shall be numerically satisfied in general, it is necessary to choose the units in such a way that it is satisfied in one particular experiment.

Equation (2) can be generalized to hold for an arbitrary number of basic processes coupled together, so that we can write as a general postulate based on experiment,

$$\Sigma\delta A=0, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \qquad (3)$$

where the summation extends over all the kinds of work involved. This law is postulated to apply to all reversible processes, and will be termed the *work principle.* It is the basis of all equilibrium expressions used in thermodynamics, physics and chemistry.

The amount of work associated with any basic process is generally determined by coupling it reversibly to a second basic process, which is particularly simple from an experimental point of view and in which the amount of work can be varied by altering only the quantity involved. An arrangement of this kind will be termed a *work reservoir*, and a simple example consists of a weight which moves at right-angles to the earth's surface. By the work principle (3) the work δA required is numerically equal to the work δA° in the work reservoir, expressed in the conventional units adopted for this reservoir. The quantity δK is obtained on the basis of the transportability condition, using a standard state in which the change in quantity is fixed by convention. Finally, δA and δK having been determined in this way, the potential difference P_1-P_2 is obtained from equation (1). It is obvious that this method, based on the energetics of reversible processes, will lead only to the potential differences, and not to the single potentials.

In traditional thermodynamics, where work is merely or mainly of mechanical nature, it is customary to say that work is " done by " a system or " done upon " a system if in a work reservoir coupled to it the basic process is negative or positive respectively. In the present treatment of the energetic phenomena, where several kinds of work are considered and where the basic process is defined by changes in the reacting system itself, it will be more descriptive and also more logical to introduce the concepts of " loss " or " gain " of work associated with the process under consideration. In fact, we shall define δA given by equation (1) as the *loss of work* associated with the passage of δK from the initial to the final state. *Gain of work* is then represented by $-\delta A$. These definitions hold irrespective of the reversibility of the process.

The following table gives a number of conjugate potentials and quantities, and the corresponding expressions for the work.

Work.	Potential.	Quantity.
$(\Phi_1-\Phi_2)\delta m$ ……	Gravitational potential.	Mass.
$-(p_1-p_2)dv$ ……	Pressure.	Volume.
$(T_1-T_2)\delta S$……..	Temperature.	Entropy.
$(\mu_1-\mu_2)\delta n$ ……..	Chemical potential.	Amount of substance.
$(\psi_1-\psi_2)\delta\epsilon$ ……..	Electrical potential.	Electric charge.
$(\gamma_1-\gamma_2)\delta\sigma$ ……..	Surface tension.	Area.
$(k_1-k_2)\delta l$ ……..	Force.	Length.

3. *Entropy and Temperature.*

As will appear from the previous sections, it is an essential feature of the present formulation that the various basic processes and kinds of work appear in a completely parallel and symmetrical manner in all changes of a reversible nature. In the *thermal* basic process the quantity which is transported from one state to another has been termed entropy, while the potential which characterizes the states from a thermal point of view is the temperature. These definitions will not be found to conflict with classical terminology.

In order to establish the magnitude of an amount of entropy which is transported or received, we shall consider a system made up of arbitrary quantities which is homogeneous with respect to these quantities and with respect to the corresponding potentials. We suppose that this system can be brought into thermal communication with its surroundings, but is otherwise isolated with respect to quantities. Such a system will be termed a calorimeter. Since it can only alter its properties by receiving thermal quantity, the amount of this quantity received can be defined by the resulting change in the properties of the calorimeter; for example, in its temperature as registered on a conventional thermometer. The amount of entropy received can then be taken as proportional to the observed rise in temperature (supposed small). Such a calorimeter can be employed as a *standard system*, a conventionally fixed proportionality factor being introduced for determining the entropy in terms of the temperature change.

By this method it is possible to determine entropies under given conditions, in particular at constant temperature. The comparison of entropies outside these standard conditions can be effected by means of the basic process, entropy being defined as *reversibly transportable* in conformity with the considerations advanced above. If an arbitrary system is in reversible thermal communication with the standard calorimeter, the entropy which it receives will be identical with the entropy given up by the standard calorimeter. In this way it is possible to compare amounts of entropy experimentally under arbitrary conditions.

The definition given here for the entropy received under arbitrary conditions will be unequivocal (and therefore acceptable) provided that the ratio between the amounts of entropy received in the initial and final states of a reversible thermal process is found to be independent of the path by which the process takes place, the amounts of entropy being measured by the method given for constant conditions, based upon conventional temperature changes. That the definition is in fact unequivocal follows from our fundamental assumption that a reversible process is a basic process in which quantities are transported, but not produced or consumed.

We have already seen how a potential difference can be determined in general on the basis of the work principle. This principle can be applied to determine the scale of thermal potential, or temperature, by inserting the values for the thermal work and quantity in the equation

$$\delta A=(T_1-T_2)\delta S. \qquad \dots\dots\dots \qquad (4)$$

The scale obtained in this way will be termed the *energetic temperature scale*. Just as for potentials in general, it will be seen that this method leads only to a scale for temperature *differences* or temperatures reckoned from a conventional zero, and does not give an " absolute " value of the temperature. As we shall see later, the energetic temperature scale is in conformity with the Kelvin scale.

Equation (4), expressing the work of a reversible thermal basic process, is of fundamental importance in thermodynamics. It gives a direct and general expression for the energetic behaviour of a " heat engine."

It is of interest to note that a system of the type described above as a calorimeter can be employed equally well for measuring any other energetic quantity, provided that this quantity only is allowed to enter or leave the standard system. It might therefore be described in general as a "quantometer." The measurement in arbitrary systems can be carried out either under given conditions or under varying conditions, just as in the case of the entropy measurement.

4. *The Concept of Heat and Energetic Heat Evolution.*

According to the basic point of view developed in the preceding sections, reversible natural phenomena can be described and interpreted in all respects in terms of energetic quantities and potentials. In this field of energetics there is, therefore, no need for any other kind of concept.

In agreement with this statement the *concept of heat* does not appear as a necessary or convenient element in the description of natural phenomena until we turn to the consideration of *irreversible* processes. In these processes we meet with new phenomena which differ fundamentally from all those so far introduced, and which need a special interpretation to render them comprehensible from an energetic point of view.

We shall begin by defining an amount of heat which is supplied to an energetic system from its surroundings under constant conditions, *i. e.*, under the conditions which we have prescribed for the standard calorimeter arrangement. Under these conditions the definition of heat is identical with the definition of entropy. This means that the amount of heat received by a calorimeter at constant temperature is proportional to the amount of entropy simultaneously received. In other words, it is not possible under these conditions to distinguish between the heat received and the entropy received.

The peculiarity of the heat concept only appears when the problem arises of comparing amounts of heat at different temperatures. While entropy is reversibly transportable, we shall define heat in general by its *irreversible transportability*, *i. e.*, if two systems are in direct thermal contact the amount of heat which leaves the one system will be by definition equal to the amount of heat received by the other.

We shall see later, on the basis of the energy principle of thermodynamics, that it cannot lead to any inconsistency if we assume that the proportionality factor between entropy and heat (defined as above) is uniquely determined by the temperature. It is therefore permissible to use this proportionality factor as a measure of the temperature by writing

$$\delta Q = T\delta S. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The scale obtained in this way will be termed the *thermal temperature scale*. When applied to thermal communication between two calorimeters of temperatures T_1 and T_2, equation (5) gives the relation

$$\frac{T_1}{T_2} = \left[\frac{\delta Q_1}{\delta Q_2}\right]_{\text{rev.}} = \left[\frac{\delta S_2}{\delta S_1}\right]_{\text{irr.}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

so that the ratio of the temperatures of the calorimeters is equal to the ratio of the amounts of heat for reversible thermal communication, and equal to the inverse ratio of the amounts of entropy for irreversible communication.

Equation (5) makes it possible to determine an absolute value for the temperature. This possibility (depending on the *production* of entropy in irreversible processes) gives the temperature a special position among energetic potentials, since it has already been shown that in general only the difference $P_1 - P_2$ is accessible experimentally.

The first of the equations expressed by (6) shows that the therma temperature scale coincides with the Kelvin scale when heat is defined as irreversibly transportable. In order to compare the thermal scale with the energetic scale set up in the preceding section, we shall consider a thermal basic process for which equation (4) holds when T is the energetic temperature. If system I loses an amount of Entropy δS as a step in the transport process, then, according to equation (5), it must simultaneously lose an amount of heat $\delta Q_1 = T_1'\delta S$, where T_1' is measured on the thermal scale. In the same way it is seen that system II must receive an amount of heat $\delta Q_2 = T_2'\delta S$. The total amount of heat lost is thus given by

$$\delta Q = (T_1' - T_2')\delta S, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and comparison with equation (4) shows that the two scales will coincide provided that we can write $\delta Q = \delta A$. Their identity thus depends on the law of equivalence of heat and work, which will be considered below.

The heat concept thus possesses a unique character, depending upon the possibility of the irreversible transport of heat without other associated changes of any kind whatever. This unique character demands an interpretation in terms of the laws of nature which are valid in other fields of energetics, and for this purpose we shall first have to consider the peculiarities of irreversible processes from a general standpoint.

Classical thermodynamics assumes that certain irreversible processes are accompanied by a loss of work and an equivalent evolution of heat. This idea, which applies for example to simple mechanical and electrical processes, is the basis of the concept of energy and the energy principle in classical thermodynamics. In our presentation the principle of irreversible heat evolution and the equivalence of heat and work will be assumed to be valid for all kinds of work, it being postulated that *an irreversible basic process is in all circumstances accompanied by a net loss of work and the evolution of an equivalent amount of heat.* In the case of a reversible process the loss of work given by equation (1) is associated with an equal amount of work gained, *e. g.*, in a work reservoir coupled to the system. In an irreversible process the same expression will hence represent both the loss of work and the heat evolution, *i. e.*

$$\delta Q = (P_1 - P_2)\delta K. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The amount of heat thus produced and given by equation (8) will be termed the *energetic heat evolution.*

It follows from these considerations that if an infinitesimal basic process is carried out reversibly and irreversibly from the same initial states, then the final states in the two cases will differ infinitesimally. If the two final states are to be made identical, then an amount of heat given by equation (8) must be either supplied to the system in the reversible case, or removed from the system in the irreversible case. If the irreversible process is of finite extent, then the heat evolution is obtained by integrating (8) along the path determined by the irreversible change. The energetic heat evolution can therefore be identified in general with the amount of heat which must be supplied from the surroundings in a reversible cinematographic reproduction of the irreversible process, *i. e.*, when the irreversible process is carried out reversibly in such a way that the course of a given element of one process is identical with the course of the corresponding element of the other process.

The above considerations show that the energetic heat evolution can be identified with an amount of heat supplied from the surroundings. It is therefore clear that an irreversible process must also be accompanied by an increase of entropy, which is related to the heat evolution by the equation

$$\delta Q = T\delta S, \quad \ldots\ldots\ldots \quad (9)$$

which is of the same form as (5), or, in the case of processes of finite extent, by the equation

$$Q = \int_1^2 T\delta S. \quad \ldots\ldots\ldots \quad (10)$$

The same relation will thus hold between entropy and heat, independent of whether the thermal magnitudes in question are produced in the system or are supplied to it. It will be seen from the above that the amount of entropy in an isolated system is not necessarily constant, as in the case for all other energetic quantities, but that it increases when irreversible processes take place. However, since the entropy produced irreversibly can be quite generally replaced by entropy supplied, the entropy will be a true *function of the state of the system*, as in the case for the other energetic quantities.

Since heat is defined on the basis of its irreversible transportability, the introduction of the concept of "energetic heat evolution" will be of decisive importance for our understanding of the concept of heat itself. It is possible to interpret the process of heat conduction in harmony with the general energetic principles presented above by regarding it as an irreversible basic thermal process, *i. e.*, a transport of entropy in which heat is *developed* simultaneously on account of its irreversibility. Suppose that an amount of heat Q flows from T_1 to T_2, a steady state having been reached. Then the amount of heat passing a section of the conductor at a temperature T will be $Q=TS$, while the amount of heat produced in the temperature interval $T+dT$ to T is given by the principle of energetic heat evolution as

$$dQ = -SdT = -Qd\ln T. \quad \ldots\ldots \quad (11)$$

The amount of entropy produced in the same temperature interval is

$$dS = -S\,d\ln T. \qquad (12)$$

By integrating between the initial and final temperatures we obtain

$$\ln\frac{S_1}{S_2} = -\ln\frac{T_1}{T_2}, \qquad (13)$$

where S_1 is the amount of entropy which leaves T_1, and S_2 the amount arriving at T_2. Equation (13) can also be written

$$T_1S_1 = T_2S_2 \qquad (14)$$

or,

$$Q_1 = Q_2, \qquad (15)$$

so that this interpretation of the transport of heat agrees with reality.

A stream of entropy passing through a thermal conductor will thus produce an irreversible heat effect in the conductor, just as in the case of an electric current, and the heat developed is calculated in an exactly analogous manner in the two cases. In spite of the simplicity which characterizes the concept of heat from an experimental point of view, it thus has from an energetic point of view a composite nature, and cannot be understood on an energetic basis without closer analysis. It is only in an irreversible thermal process that the quantity transported is of the same kind as the quantity produced irreversibly, and it is of course this circumstance which has hitherto served to obscure the relationships which exist.

5. *Energy.*

The real basis for the energy concept of thermodynamics is the law of the equivalence of heat and work. As long as our considerations are confined to reversible processes, which are fully described by the validity of the work principle $\Sigma\delta A = 0$, the description of work as " energy " (or of loss of work as loss of energy) will not contribute any new concept, and cannot be attributed any real significance. It is equally true that the description of heat as " energy " will not in any way be a new characteristic of heat in processes in which heat is only transported in a reversible manner, *i. e.*, at constant temperature. If, on the other hand, our considerations are extended to include irreversible processes, in which we have assumed that work is in general converted into heat, then the numerical equivalence which is found to exist is evidence of a common concept which includes both these quantities, and for which it is natural and useful to introduce a special term. It is on this basis that the concept of *energy* in thermodynamics is derived, and the law of the conservation of energy can be maintained.

If we consider a system which is isolated from its surroundings in all respects, the energy equation for such a system can be written

$$\delta E \equiv \delta Q - \Sigma \delta A = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where $\Sigma\delta A$ is the total work lost by the system and δQ the energetic heat evolution. This equation expresses the work principle and the principle of the equivalence of heat and work. By using equations (1) and (5) for the work and the heat respectively, (16) can also be written

$$\delta E = \Sigma P dK = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

both heat and work occurring in the form expressed by (17).

If we now imagine that the closed system mentioned above is divided into two subsidiary systems communicating with one another in an arbitrary manner, then for each of these subsidiary systems we can write an equation of the form

$$dE = \Sigma P dK, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where it may now no longer be possible to divide the single terms of the form PdK into groups representing heat and work respectively. Equation (18) is therefore a general energy definition which states that in a system communicating with its surroundings in an arbitrary manner the increase of energy is given by the sum of all the particular kinds of energy communicated from outside the system, each of these being given by the product of a quantity and the corresponding potential in the system itself.

Equation (18) is a convenient energy expression and is in agreement with, for example, Gibbs' formulation. However, it is clear that it is essentially a definition, and does not express any conceptual law or any verifiable principle.

If the quantities concerned in the system are only volume, entropy, and amount of matter, then (18) may be integrated (as in Gibbs' treatment), giving

$$E = \Sigma PK. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Differentiation and combination with (18) then gives

$$\Sigma K dP = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

Thus, whereas energy changes of the form PdK go to make up the total energy change, energy changes of the form KdP will have no effect on the total energy, since according to (20) their sum automatically vanishes.

6. *Heat and Energy in Traditional Thermodynamics.*

In this section we shall give a treatment of the concepts of heat and energy in traditional thermodynamics, with special reference to the way in which these concepts differ in principle from those advanced in the present paper.

As regards the heat concept itself, it is usually introduced into thermodynamics without any real or satisfactory definition, and without any attempt to interpret its peculiarities from a general energetic standpoint. When dealing with the supply or transport of heat, it is usually treated in conformity with the definition given in section 4 of the present paper. However, in the question of heat *evolution* and the general *relation between heat and work*, the classical interpretation differs vitally from the view taken in the present treatment. Thus we have supposed that in reversible processes interconversion takes place between different kinds of work in accordance with the work principle, no heat being consumed or produced: on the other hand, according to the classical view these processes may be accompanied by either the " evolution " or the " absorption " of heat. We shall now proceed to test the validity of this supposition.

It must first be remembered that the description of reversible processes based upon the work principle is characterized by the complete symmetry with which all the energetic quantities and potentials appear in both the general and the particular equations. On this basis, therefore, any asymmetry which is introduced into the treatment of reversible processes must be based upon a lack of symmetry in the problem itself. Thus, when the traditional treatment finds it necessary to introduce a special form of energy in addition to thermal work, which appears in reversible processes cognately with the other forms of work, it is necessary to investigate very closely the basis for such a striking assumption.

This investigation is rendered more difficult by the fact that the theory of the reversible heat phenomena is not presented anywhere in classical thermodynamics in even a moderately consistent form. No statement is given of the general principles upon which it is possible to demonstrate the occurrence of a reversible heat change, or to measure its magnitude. However, a study of the cases in which the reality of the heat phenomenon is considered to be specially well established gives the impression that, according to the traditional point of view, it is the *temperature change* in the system which serves as a criterion of the occurrence of heat changes both in reversible and irreversible processes, and that the sign of this heat change is determined by the sign of the temperature change. Similarly, the magnitude of the amount of heat produced or consumed appears to be calculated as the amount of heat which must be removed from or supplied to the system in order that the process in question shall take place isothermally. It is on account of this point of view that we shall describe the heat changes of the traditional treatment as *thermometric heat changes*.

It will at once be seen that this concept is in principle open to the objection mentioned above, based upon the lack of symmetry thus introduced in the treatment. Thus, it may be asserted that a temperature change in a reversible process can in no way be allotted a more fundamental significance than a change in other energetic potentials occurring in the system, *e. g.*, pressure or chemical potential. It is therefore arbitrary and irrational to assume the production of a new form of energy on the basis of an increase of temperature without making corresponding assumptions for changes in pressure or chemical potential.

This argument is of such decisive importance that even by itself it would appear to provide sufficient grounds for dismissing the traditional assertion of heat changes in reversible phenomena. The same necessity is apparent when we consider the condition demanded by the work principle, that the loss of work in a closed reversible system must be zero. If heat were produced or consumed in a closed system during a reversible process, this would conflict with the equivalence principle and the energy principle, since it is not possible to point out any equivalent for the heat produced or consumed.

Even without considering the work principle in its most general form, the same conclusion can be reached by considering a simple case in which there can be no doubt about the various work terms involved. For example, suppose that the system in question is a gas plus a vacuum, and that the two volumes are changed reversibly by the displacement of a piston attached to a work reservoir. The loss of work $p\delta v$ in the system will be associated with a gain of work $-\delta A$ in the reservoir. These two amounts of work are found to be identical when measured in the same mechanical units ; hence, according to the equivalence principle, we can exclude the possibility that heat is produced or consumed, since also according to classical ideas heat is a form of energy and cannot be produced without an equivalent loss elsewhere.

The above process carried out with *ideal gases* has played an important part in the theory of reversible heat phenomena and of the mutual transformation of heat and work, since it has been regarded as a particularly lucid example for illustrating the transition from heat to work and *vice versa.* Thus if an ideal gas in thermal communication with a large heat reservoir is compressed, the amount of heat received by the reservoir will be equal to the amount of heat supplied, and the compressed gas can now expand irreversibly without any change of temperature, the system thus returning to its original state. At the same time, the actual net result will be observed as a caloric phenomenon, consisting of the conversion of the work supplied into an equivalent amount of heat. According to classical thermodynamics, it is the *reversible* compression which causes the loss of external work and the formation of an equivalent amount of heat which is transferred to the reservoir, while the subsequent irreversible expansion is an inactive additional phenomenon, which takes place without "energy transformation." On the other hand, according to the present treatment the loss of work and the production of an equivalent amount of heat will of course be associated with the irreversible part of the process, while the reversible part (in which the total loss of work is zero) is only responsible for the increase of temperature, which occurs simultaneously with changes in the other energetic potentials. and which makes possible the neutral transfer of a finite amount of heat to the reservoir.

The specially simple behaviour which characterizes ideal gases from a physical point of view makes them to some extent dangerous objects for energetic investigations, which are designed to discover general regularities and fundamental principles. The consideration of special cases is often an unfortunate method of investigating general principles, and the above traditional interpretation of the process of expansion is a noteworthy example of this. The obscurity of the subject is also undoubtedly due in part to the introduction of arguments from molecular theory into energetic considerations, the kinetic energy of the molecules being identified with "heat," and the latter term being used more as an abstract qualitative designation than as a magnitude or quantity which is sharply defined from an energetic standpoint.

We have so far dealt particularly with reversible processes in which the classical treatment considers heat to be produced. In other processes heat is supposed to be consumed, an equivalent amount of work being simultaneously produced. These processes have played an equally important part in classical thermodynamics ; in fact, the formulation of the so-called second law of thermodynamics is in reality based on a consideration of processes of this kind. The idea which has taken root in this field supposes that the transformation of heat to work cannot take place as an *isolated* process, but that such a transformation can be carried out if it is associated with concurrent changes of another kind. These simultaneous processes are described as "compensations."

It is clear that such a possibility is in direct contradiction to the ideas advanced in this paper, according to which a gain of work can only be produced at the cost of an equivalent loss of work. In order to illustrate this contradiction it will be convenient to consider two typical examples. Authoritative accounts of thermodynamics (*e.g.* by Planck) contain the statement that the isothermal and reversible expansion of a perfect gas is a process in which complete conversion of heat into work takes place. This statement is based upon the fact that the work stored up in the work reservoir R_A, coupled to the system, is equal to the heat simultaneously supplied from an external heat reservoir R_Q to the system gas+vacuum, and that apart from this no "energy transformations" whatever take place in the system.

It is, however, easy to see that this view is based on a rather superficial analysis of the phenomenon in question. Thus, it is maintained that it is the amount of heat supplied from the external reservoir R_Q which is converted to the work stored up in the external reservoir R_A, and that apart from this no energy transformations take place. However, this interpretation completely overlooks the fact that the system gas+vacuum has also been changed during the expansion, and that this change is identical with the change undergone by the system when the expansion in question takes place irreversibly without any connexion with the surroundings. If, therefore, the classical view is to be logically maintained, it must be assumed that the *reversible* expansion of a gas must be considered as the sum of two processes, of which one is *irreversible* and the other *impossible*. It is, however, irrational to divide up a concrete phenomenon in this way, and further evidence will be produced later to show that this assumption is not acceptable as a serviceable interpretation.

As our second example we shall consider the *heat engine*. The theory of the heat engine constitutes the branch of thermodynamics in which the idea of the transformation of heat to work has played the most prominent part, and in which the supposed energy transformation is regarded as a particularly palpable reality. It was in this field that Carnot advanced his penetrating idea without being aware of (or at least without making use of) what is commonly known as the law of conservation of energy. The combination of this law with Carnot's principle is due to Clausius. In this way thermodynamics reached a form in which the basic ideas have remained unassailed until the present day.

It is well known that Clausius' solution depends upon splitting the heat energy available for the production of work into two parts with different functions. The work obtained from the heat engine is equated to an amount of heat simultaneously absorbed, so that the first law is obeyed. Since heat cannot be directly converted into work without "compensation," the necessary compensation is associated with the conversion of heat by the passage of heat from a higher to a lower temperature. However, in Clausius' treatment this last process is not considered to be the actual cause of the production of work, as is the case with Carnot. In fact, it may be said that the introduction of the first law and the recognition of the numerical equivalence has served to a great extent to obscure what Carnot supposed to be the basis of the transformation of heat into "motive power." Further, it is in definite contrast to the factor which in the present treatment is assumed to determine all energy transformations, *i. e.*, the potential difference present in the functioning system. Although Clausius' system leads to an unimpeachable basis for thermodynamic calculations and the correlation of energetic phenomena, it does not seem that it can be regarded in principle as a step forward in the conception and understanding of the fundamental phenomena of thermodynamics.

The classical view can be expressed schematically as

$$\left.\begin{array}{r}(Q_1-Q_2)_1\rightarrow A,\\ (Q_2)_1\rightarrow(Q_2)_2\end{array}\right\}, \quad \ldots\ldots\ldots \quad (21)$$

or

$$\left.\begin{array}{r}(Q_1-Q_2)_2\rightarrow A,\\ (Q_1)_1\rightarrow(Q_1)_2\end{array}\right\}, \quad \ldots\ldots\ldots \quad (22)$$

or

$$\left.\begin{array}{r}(Q_1-Q_2)_3\rightarrow A,\\ (Q_1)_1\rightarrow(Q_1)_3,\\ (Q_2)_3\rightarrow(Q_2)_2\end{array}\right\}, \quad \ldots\ldots\ldots \quad (23)$$

where the index outside the parentheses refers to the temperatures T_1, T_2 and T_3 at which the amounts of heat act. T_1 and T_2 are the temperatures of two external heat reservoirs with which the system communicates during its action, and T_3 is the temperature of a heat reservoir belonging to the system. The net amount of heat received by this last reservoir is zero, so that it can be of arbitrary capacity. These interpretations are formally different, but are completely identical from a phenomenological point of view.

In considering Clausius' interpretation critically it must first be emphasized that, from a general point of view, it seems incomprehensible that a transformation of heat which is recognized to be impossible as an isolated process should become possible merely by associating with it certain subsidiary processes which have no genetic relation to the work produced. Clausius makes no attempt to justify this state of affairs, being concerned solely with the quantitative equivalence. However, it is not all that the matter is incomprehensible from a general point of view. If we investigate more closely the real nature of the process introduced as "compensation," *i. e.*, the passage of an amount of heat from one temperature to another, it will be realized that it is not justifiable to assume such a compensation in a reversible process of the kind involved in the working of a heat engine, since (as shown in section 4) the transport of heat involves the irreversible production of heat, and can therefore never form a rational component of a reversible process. The undoubted misunderstanding which underlies Clausius' interpretation is closely connected with the concept of heat as a generally transportable quantity, or of the transport of heat as what has been termed a basic energetic process in this paper. In Clausius' considerations the unrealizable compensation process is associated with an equally unrealizable main process, and it is only for this reason that the total process which he arrives at has a place in the world of reality. It is therefore impossible to attribute any physical reality to the interpretation employed by Clausius or to the consequent assertion that it is possible to transform heat into work.

The features which characterize the classical interpretation of the special examples discussed here are present in general in all attempts to explain the transformation of heat to work. As long as the intrinsically impossible transformation

$$Q_1 \rightarrow A_1$$

is accepted as a link in a given natural process, it will always be necessary to introduce as " compensation " a process of the form

$$A_2 \rightarrow Q_2,$$

where A_1 and A_2 are equal amounts of different kinds of work, and Q_1 and Q_2 are equal amounts of heat respectively absorbed and evolved at the same temperature. However, it may be said quite generally of this type of explanation that it will permit *any process whatever* to appear as a component of an actual total process, since no matter what may be the form of the postulated partial process it will always be possible to find a compensating partial process such that the two together produce the actual process. The complete arbitrariness thus introduced into thermodynamic methods would deprive such a principle of any real significance.

It is a decisive point in the above considerations that *a temperature change cannot in any way be regarded as a general sign of a caloric process taking place in the system*, if we understand by this the production or consumption of an amount of heat to which there corresponds an equivalent amount of work. There is no doubt that a " thermometric heat change " takes place, since, according to the above, this is *defined* by the temperature change. This is, however, purely a question of terminology, and to attribute to it the character of heat in an energetic sense would lead to conflict with the energy principle. It is only when the system is a " calorimeter " that there is a simple relation between the temperature change and the amount of heat, as in the method used for actually defining heat. It also follows that a temperature change cannot be used for measuring the heat which is produced in irreversible processes and which is equivalent to the loss of work, since the temperature change is only partly determined by the irreversibility. In this case the amount of heat must be measured in a different way, as described in section 4.

7. *The Fundamental Principles of Energetics.*

The preceding sections deal with the general laws for natural phenomena regarded from a macroscopic standpoint, and thus represent what might be termed the fundamental energetic principles. In the present section these principles will be collected and formulated precisely.

On the basis of earlier considerations we shall postulate the following law as being universally valid :—

Natural processes which take place spontaneously are accompanied by a positive total loss of work and a positive heat evolution. These two quantities become zero in the limiting case of a reversible process, and are numerically equivalent for all kinds of changes.

The content of this law can be expressed briefly in the following form, which also serves to characterize processes which cannot be carried out, and which we shall term imaginary :—

$$\Sigma A = Q \begin{matrix} > \\ = \\ < \end{matrix} 0 \quad \left.\begin{matrix} 1.\ \text{irrev.} \\ 2.\ \text{rev.} \\ 3.\ \text{imag.} \end{matrix}\right\} . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

The various laws contained in this expression can be expressed separately as follows. In the first place,

$$\Sigma A = Q, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

which is the *thermodynamic energy* or *equivalence principle.* Secondly,

$$\Sigma A \equiv Q > 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

a law which we shall term the *heat principle,* and which states that all irreversible processes are accompanied by loss of work (and evolution of heat). Thirdly,

$$\Sigma A \equiv Q = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

which states that in all reversible processes the total loss of work (and the heat evolution) is zero. This law has already been termed the *work principle.*

Finally, equation (24) contains an expression for impossible or imaginary processes,

$$\Sigma A \equiv Q < 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

in which the validity of the thermodynamic energy principle has been formally retained. In these processes both the loss of work and the evolution of heat are negative ; *i. e.*, a conversion of heat into work takes place. The present system of thermodynamic principles is based upon the realization that such processes are impossible.

In comparing these principles with the laws of traditional thermodynamics, it must be remembered that the traditional concepts of both work and heat evolution differ in an essential manner from the corresponding concepts in the present treatment. It will therefore be found that the two traditional laws of thermodynamics cannot be compared directly with any of the laws represented by equations (25) to (28).

However, we shall not give here an account of the considerable differences which characterize the basic principles of the two systems, partly because it would require a comprehensive treatment even to describe the various formulations under which these appear in the classical system of laws. On this question reference should be made to the two original papers, of which the present paper is an extract.

Only two points will be emphasized here. One of the traditional expressions of the second law of thermodynamics is :—

$$\triangle S \geqq 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

If we remember the meaning of the concept of energetic heat evolution, it will be seen that this expression is included in the heat principle and the work principle in the form

$$Q \geqq 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

since heat and entropy are connected by the general relation $\delta Q = T\delta S$. In classical thermodynamics it has been considered an essential point in (29) that S is a function only of the state of the system, and $\triangle S$ therefore a function only of the initial and final states of the system. This is, of course, correct if the final state includes those parts of the reacting system which receive the irreversibly produced entropy.

If, on the other hand, "the process" refers only to the sum of the *basic processes* which it involves, so that the irreversibly produced entropy is collected in such a way that it does not form part of the final system, then in a process of this kind it will be the amount of heat, and not the entropy, which is a function only of the initial and final states. The two cases can be represented schematically by

$$[A \rightarrow B + \triangle S]$$
$$[A \rightarrow B] + \triangle S,$$

where the brackets include the reacting system. In the first case it is the entropy, and in the second case the heat evolution, which is determined uniquely as a function of the initial and final states of the process.

As a second point, we shall emphasize the special position of the work principle. Without the introduction of special assumptions this principle cannot be derived from any of the other principles, or from the laws of traditional thermodynamics. The work principle is the rational expression of the peculiarities of reversible processes, and for this reason forms the basis of all the equlibrium expressions of thermodynamics. It is therefore particularly clear that it cannot be derived from the second law (although traditional thermodynamics usually makes this law the basis of equilibrium expressions) as long as we regard this law as being expressed by an inequality such as (29) rather than by an equality. The special character which must be assigned to the work principle appears when we consider reversible processes as the boundary case between irreversible processes for which

$$A \rightarrow Q,$$

and imaginary processes, for which

$$Q \rightarrow A,$$

contact between these two groups being only possible for $Q=0$. Under these conditions the directional nature of the two groups of processes disappears, and the two expressions degenerate to

$$A=0,$$

which is the work principle.

8. *Summary.*

This section contains a collection of the most important of the results concerning the relation between heat and work which have been obtained in this paper.

1. It is possible to construct a general system of energetic principles on the basis of the concepts *quantity* and *potential*. A *basic process* in energetics consists of a transport of quantities. All reversible natural processes consist exclusively of such basic processes.

2. Every basic process involves a *loss or work*, which is written in the work equation as the product of the quantity transported and the potential difference between the two states between which the transport takes place, *i. e.*, $\delta A=(P_1-P_2)\delta K$. For reversible processes the total work is zero: $\Sigma\delta A=0$. This law is termed the *work principle*.

3. The basic thermal process is a transport of entropy between different temperatures. Entropy is the thermal quantity, and is defined in general on the basis of its *reversible transportability*.

4. When a system receives entropy it also receives by definition an amount of heat which is proportional to the entropy. Heat is defined in general on the basis of its *irreversible transportability*. Heat and entropy supplied are connected by the relation $\delta Q=T\delta S$, which defines temperatures on the "thermal" scale. If the equivalence of heat and work is assumed this scale is identical with the "energetic" temperature scale defined by $\delta A=(T_1-T_2)\delta S$. This is also identical with the Kelvin scale.

5. Irreversible processes are accompanied by loss of work and evolution of heat, $\Sigma\delta A>0$, $\delta Q>0$. This law is called the *heat principle*. Simultaneously entropy is produced, which is related to heat in the same way as entropy supplied, *i. e.*, by the relation $\delta Q=T\delta S$.

6. The loss of work in an irreversible process is given by the work equation. It can be determined as the external work obtained when the process is reproduced in a reversible manner. The excess of heat which

must be supplied from the surroundings in carrying out this reversible reproduction is known as the *energetic heat evolution.*

7. The loss of work in an irreversible process is proportional to the energetic heat evolution. On the basis of this result the *equivalence principle* $\Sigma\delta A=\delta Q$ is postulated.

8. The special position occupied by heat in energetics is comprehensible on the basis of an energetic interpretation of the *process of heat conduction.* If this is regarded as an irreversible basic process accompanied by the production of entropy and heat, then agreement with experience is obtained.

9. It is not possible in general to measure or define the absolute values of the energetic potentials on an energetic basis. This is only possible in the case of the thermal potential, temperature, by applying the equivalence principle to an irreversible process in which an amount of heat of the form $T\delta S$ is produced.

10. Caloric phenomena are *confined to irreversible processes.* In traditional thermodynamics it is supposed that reversible processes may be accompanied by the evolution or absorption of heat, the appearance of which is shown by changes of temperature, but this supposition rests on an insufficient analysis of the phenomena in question. Heat changes of this kind cannot be reconciled with the validity of the equivalence principle and the work principle.

11. *The conversion of heat into work is a process which cannot be carried out under any circumstances.* Traditional thermodynamics represents the reversible expansion of a perfect gas as an example of such a conversion, but such a view fails to take into account the loss of work represented by the change in volume. If this work is taken into consideration, the assumption that heat is absorbed would not be compatible with the equivalence principle and the work principle.

12. The action of the heat engine can be described as a *basic thermal process* in which entropy is transferred from a higher to a lower temperature, coupled with a mechanical process. The expression for the work of such processes, $\delta A=(T_1-T_2)\delta S$, is in quantitative agreement with Clausius' expression, which is verified by experiment.

13. Clausius' traditional interpretation of the action of a heat engine assumes that the transformation of heat into work is accompanied by a compensating transport of heat. However, this cannot be regarded as a satisfactory assumption, since the transport of heat, which is *irreversible,* cannot enter as a component of a *reversible* process. The introduction of "compensations," as in the classical treatment, must be rejected as a principle of scientific interpretation on account of the arbitrariness which it involves and the imaginary processes which it employs.

14. The concept of heat changes used in traditional thermodynamics applies both to reversible and to irreversible processes, and is based upon

temperature changes in the system ; it is therefore described as *thermometric heat change* in the present treatment. It can be defined as an amount of heat, but *there is no work equivalent for the evolution or the absorption of such an amount of heat,* as is demanded by the equivalence principle and the energy principle.

15. *The concept of energy* in thermodynamics is based upon the equivalence of heat and work. The real content of the *energy principle* is contained in the equivalence principle. *The general energy definition* $dE = \Sigma P dK$ and other more extended formulations of the energy principle have only a formal significance.

16. *The fundamental principles of energetics, comprising the work principle, the heat principle and the equivalence principle, can be summarized by the expression* $\Sigma \delta A = \delta Q \geqq 0$. The real content of the first law of thermodynamics is covered by the *equivalence principle.* If the second law is expressed by the inequality $\delta S > 0$, then its content is equivalent to that of the *heat principle.* The *work principle* is valid for all reversible processes and is the basis of all equilibrium equations in energetics and thermodynamics. It does not follow directly from the two laws of thermodynamics as formulated here, but is obtained as a consequence of the fact that reversible processes occupy a boundary position between irreversible processes on the one hand, and on the other hand those imaginary processes which are schematically permissible, but which cannot be carried out in practice.

University Institute of Physical Chemistry,
Copenhagen.

XLVI. *Gravity and the Rotation of the Earth.*

By C. G. PENDSE, M.A., Ph.D.*

[Received June 10, 1939.]

THE connexion between the acceleration due to gravity ("g") and the rotation of the earth is generally not clearly brought out in the published literature. In particular the dependence of g on the terrestrial frame of reference to which it is related and the way in which, in practice, it is made to depend on the origin of the frame are not clearly indicated. The object of this note is to show the successive steps by which, starting from the law of gravitation and the "fixed" axes for the solar system, a "terrestrial frame," *i. e.*, a set of axes fixed relative to the earth, comes into prominence and the nature of the relation of the frame to the "acceleration due to gravity."

We can classify the problems of Mechanics under two headings: (1) Celestial Mechanics and (2) Terrestrial Mechanics; and the question of gravity arises in the second section.

In problems of Celestial Mechanics it is generally not necessary to consider "contact forces"; all the forces are "body forces," *i. e.*, gravitational forces, and the directions of the axes of reference are "fixed," the origin being some suitable point, in the case of the solar system its "centre of mass." (In practice it is necessary to transfer the origin to one of the bodies in question and to transform the fundamental equations accordingly.)

In problems of Terrestrial Mechanics, however, the observations refer to axes whose directions are fixed relative to the earth and whose origin is either on or above and near the surface of the earth; it is not convenient, and often also not possible, to state the position of an object relative to axes through the centre of the earth and pointing towards the "fixed" stars. It is also necessary to consider, in some problems, "contact forces."

For the sake of simplicity we consider the motion of a particle relative to the earth. The forces on it will be :—(1) The gravitation of the earth; (2) the gravitation of other particles (or bodies) not forming part of the earth but in its neighbourhood; (3) the contact and other interaction forces, if any ; and (4) the gravitational forces due to the members of the solar system outside the earth. The forces on the earth * will be :—(1) The gravitation of the particle ; (2) the gravitation of other particles not forming part of the earth but in its neighbourhood ; and (3) the gravitational forces due to the other members of the solar system.

---

* Communicated by the Author.

The axes of reference will be through the " centre of mass " of the whole solar system and will point towards the " fixed " stars, it being assumed that the forces due to bodies outside the solar system can be ignored.

Let G be the " centre of mass " of the solar system, C the centre of the earth, which we assume is also its " centre of mass," and P the particle which is supposed to move in the neighbourhood of the earth.

For the differentiation of vectors with respect to time let $\frac{d}{dt}$ and the dot (.) denote differentiations in relation to directions fixed with reference to the " fixed " stars and directions fixed with reference to the earth respectively. Let

$\vec{I_P}$=intensity of the gravitational force at P due to the earth,

$\vec{I_E}$=intensity of the gravitational force on the earth due to P,

$\vec{F_P}$=intensity of the gravitational force at P due to the members of the solar system outside the earth,

$\vec{F_E}$=intensity of the gravitional force on the earth due to the members of the solar system outside the earth,

$\vec{L_P}$=intensity of the gravitational force at P due to other particles not forming part of the earth, but near it,

$\vec{L_E}$=intensity of the gravitational force on the earth due to other particles not forming part of the earth, but near it,

and $\vec{T_P}$=the contact and other interaction forces, if any, on the particle P ;

---

* No contact forces on the earth are introduced owing to the following reason. The earth and the particles in contact with it are supposed to form one body in so far as the motion of the earth as a whole is concerned ; consequently the contact forces like those due to friction will neutralise one another. In the case of the particle under consideration, however, the contact and other interaction forces may be due to the actions of other particles through some kind of constraint, etc. It is possible, however, that the particle in question may be in contact with the earth and may thus have a contact force from that source. While this latter contact force may be significant for the particle, it will not have an appreciable influence on the motion of the earth as a whole ; moreover, in that case, the earth together with the particle in question would form one system during the period of contact and the motion of the component would be considered. At any rate the question of one particle would not affect the earth to an appreciable extent.

let M and m be the masses of the earth and the particle respectively; and let $\vec{R}_P = \vec{I}_P - \vec{I}_E$.

The equations of motion, relative to the "fixed" axes through G, are :—

$$M\frac{d^2\overrightarrow{GC}}{dt^2} = M\vec{I}_E + M\vec{F}_E + M\vec{L}_E,$$

and
$$m\frac{d^2\overrightarrow{GP}}{dt^2} = m\vec{I}_P + m\vec{F}_P + m\vec{L}_P + \vec{T},$$

i. e.,
$$\frac{d^2\overrightarrow{GP}}{dt^2} = \vec{I}_P + \vec{F}_P + \vec{L}_P + \frac{\vec{T}}{m}, \quad \ldots\ldots \quad (1)$$

and
$$\frac{d^2\overrightarrow{GC}}{dt^2} = \vec{I}_E + \vec{F}_E + \vec{L}_E. \quad \ldots\ldots \quad (2)$$

From equations (1) and (2) we obtain

$$\frac{d^2}{dt^2}(\overrightarrow{GP} - \overrightarrow{GC}) = \vec{I}_P - \vec{I}_E + \vec{F}_P - \vec{F}_E + \vec{L}_P - \vec{L}_E + \frac{\vec{T}}{m},$$

i. e.,
$$\frac{d^2\overrightarrow{CP}}{dt^2} = \vec{R}_P + \vec{F}_P - \vec{F}_E + \vec{L}_P - \vec{L}_E + \frac{\vec{T}}{m}. \quad \ldots\ldots \quad (3)$$

$\vec{R}_P$ is the intensity of the "effective" gravitational force due to the earth on P when the motion is referred to C.

Since the particle moves near the surface of the earth, $\vec{F}_P$ and $\vec{F}_E$ differ but slightly from one another. Furthermore, $\vec{L}_P$ and $\vec{L}_E$ are small in comparison with $\vec{I}_P$. Therefore, $\vec{F}_P - \vec{F}_E$ and $\vec{L}_P - \vec{L}_E$ can be ignored in comparison with $\vec{R}_P$ and $\frac{\vec{T}}{m}$. Hence, we have

$$\frac{d^2\overrightarrow{CP}}{dt^2} = \vec{R}_P + \frac{\vec{T}}{m}, \quad \ldots\ldots \quad (4)$$

as a very good first approximation *.

---

* Some explanation is desirable as to why $\vec{I}_E$ is retained in $\vec{R}_P$ while $\vec{L}_P$ and $\vec{L}_E$ are ignored. To bring out the analogy between the earth plus the particles around it with the sun and the solar system, $\vec{I}_E$ has been kept while the other small quantities have been discarded. In considering the motion of a planet relative to the sun, the perturbations due to the other bodies (planets, etc.) are ignored in the first approximation and the planetary orbits are obtained. Furthermore, if the masses of the planets are ignored in comparison with that of the sun, Kepler's third law is obtained. In the present case it is pointed out below that effectively $\vec{R}_P$ is determined by the position relative to the earth of, and is independent of the mass of, the particle P.

Let Ω be the angular velocity, assumed to remain constant, of the earth relative to the "fixed" directions; and let O be a point *fixed relative to axes fixed in the earth* and either on or above or below the surface. Now, by Coriolis' theorem, if $\vec{A}$ is any vector,

$$\frac{d\vec{A}}{dt}=\dot{\vec{A}}+[\vec{\Omega},\ \vec{A}].$$

Hence,

$$\frac{d\overrightarrow{CO}}{dt}=[\vec{\Omega},\ \overrightarrow{CO}],$$

$$\frac{d^2\overrightarrow{CO}}{dt^2}=[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CO}]],$$

$$\frac{d\overrightarrow{CP}}{dt}=\frac{d\overrightarrow{CO}}{dt}+\frac{d\overrightarrow{OP}}{dt}=\frac{d\overrightarrow{CO}}{dt}+\dot{\overrightarrow{OP}}+[\vec{\Omega},\ \overrightarrow{OP}],$$

and

$$\frac{d^2\overrightarrow{CP}}{dt^2}=\frac{d^2\overrightarrow{CO}}{dt^2}+\frac{d\dot{\overrightarrow{OP}}}{dt}+\left[\vec{\Omega},\ \frac{d\overrightarrow{OP}}{dt}\right],\ \text{since } \vec{\Omega} \text{ is constant,}$$

$$=\ddot{\overrightarrow{OP}}+2[\vec{\Omega},\ \dot{\overrightarrow{OP}}]+[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{OP}]]+[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CO}]].$$

Equation (4) becomes

$$\ddot{\overrightarrow{OP}}+2[\vec{\Omega},\ \overrightarrow{OP}]+[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{OP}]]+[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CO}]]=\vec{R}_P+\frac{\vec{T}}{m},\quad . \quad . \quad (5)$$

i. e.,

$$\ddot{\overrightarrow{OP}}+2[\vec{\Omega},\ \dot{\overrightarrow{OP}}]+[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{OP}]]=\vec{R}_P-[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CO}]]+\frac{\vec{T}}{m},\quad . \quad . \quad (6)$$

which is equivalent to the equation

$$\ddot{\overrightarrow{OP}}+2[\vec{\Omega},\ \dot{\overrightarrow{OP}}]=\vec{R}_P-[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CO}]]-[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{OP}]]+\frac{T}{m},$$

$$=\vec{R}_P-[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CP}]]+\frac{\vec{T}}{m}.\quad . \quad . \quad . \quad . \quad . \quad . \quad (6')$$

We now confine our attention to the case in which the directions of the axes of reference are fixed relative to the earth.

It is to be noted that equation (6) is general and is valid for all positions of O—within, on or above the surface of the earth. Since $\dot{\overrightarrow{CO}}=0$, the motion and the trajectory described by equation (6) are intrinsically the same for all positions of O. The apparent importance of the position of O comes in only when the dissection of $[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CP}]]$, which is independent of the choice of O, into its components $[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{CO}]]$ and $[\vec{\Omega},\ [\vec{\Omega},\ \overrightarrow{OP}]]$

becomes a matter of *practical importance.* Incidentally, $\vec{R}_P$ and $\vec{R}_P-[\vec{\Omega}, [\vec{\Omega}, \overrightarrow{CP}]]$ are independent of the choice of O.

Since the masses of the particles, whose motions near the earth are to be considered, are very much smaller than that of the earth, $\vec{I}_E$ is negligible in comparison with $\vec{I}_P$; consequently, to all intents and purposes, $\vec{R}_P$, which equals $\vec{I}_P-\vec{I}_E$, is independent of the mass of the particle P and is determined only by the position of P relative to the earth.

When the particle moves freely, *i. e.*, when there are no significant forces other than the gravitation of the earth, $\vec{T}=0$, and the initial acceleration of the particle moving from rest at P is equal to $\vec{R}_P-[\vec{\Omega}, [\vec{\Omega}, \overrightarrow{CP}]]$ which is the same for all particles and is the "acceleration due to gravity": $[\vec{\Omega}, [\vec{\Omega}, \overrightarrow{CP}]]$ is what is usually called the "centrifugal acceleration" due to the rotation of the earth. Denoting by $\vec{g}_P$ the acceleration due to gravity *at* P, we have the equation (6′) in the form

$$\ddot{\overrightarrow{OP}}+2[\vec{\Omega}, \dot{\overrightarrow{OP}}]=\vec{g}_P+\frac{\vec{T}}{m}. \qquad (7)$$

Equation (7) is, however, not very convenient when the individual values of the acceleration due to gravity at different points are to be considered. Moreover, when O is either on the surface or is above it and at a distance small in comparison with CO—this is the usual case—and when the motion of the particle is such that OP is always much smaller than CP, $\vec{R}_P$ does not differ significantly from its value at O. Using the symbol S for the position of O as restricted above, and putting $\vec{g}_S=\vec{R}_S-[\vec{\Omega}, [\vec{\Omega}, \overrightarrow{CS}]]$, we can write the equation (6) in the form

$$\ddot{\overrightarrow{SP}}+2[\vec{\Omega}, \dot{\overrightarrow{SP}}]+[\vec{\Omega}, [\vec{\Omega}, \overrightarrow{SP}]]=\vec{g}_S+\frac{\vec{T}}{m}, \qquad (8)$$

which reduces to

$$\ddot{\overrightarrow{SP}}+2[\vec{\Omega}, \dot{\overrightarrow{SP}}]=\vec{g}_S+\frac{\vec{T}}{m}, \qquad (9)$$

if we neglect $[\vec{\Omega}, [\vec{\Omega}, \overrightarrow{SP}]]$ owing to the actually small value of $\vec{\Omega}$ and also the smallness of the ratio $SP:CS$. $\vec{g}_S$, the acceleration due to gravity near the surface at S, depends on the position of S (variation of gravity with latitude), and can be determined by experiment. The equation (8) is essentially a *local* equation in relation to the surface of the earth and can be used when the particle does not travel *widely* near or above the earth's surface; its chief merit is that the effective field force is uniform.

For many problems of terrestrial mechanics, the rotation of the earth can be altogether neglected.

In case a particle goes high up over the surface of the earth, it may be necessary to correct for the variation of gravity. If it travels widely near the surface, it may become necessary to choose different positions for S for different parts of its trajectory.

Summary.

The axes of reference for the motion of the members of the solar system, to which the law of gravitation applies, have for their origin the " centre of mass " of the system and their directions are towards the " fixed " stars. In problems of Terrestrial Mechanics, however, the observations of position refer to axes whose directions are fixed relative to the earth and whose origin is either on or above and near the surface of the earth. The successive steps by which one can obtain the equations of motion of a body (particle), valid for the frames of reference suitable for Terrestrial Mechanics, are shown. The combination of the centrifugal acceleration with the gravitational intensity due to the earth gives rise to the " acceleration due to gravity " [equation (7)], the equation being transformed into equation (8), in which the value of the quantity at the *origin* of the frame is involved. Furthermore, the mutual relations between the different terrestrial frames are pointed out.

c/o Downing College,
Cambridge.

XLVII. *On Mass and Force in Newtonian Mechanics—Addendum to "Mass I." and "Mass II."*

By C. G. Pendse, M.A., Ph.D.*

[Received July 13, 1939.]

Introduction †.

In two previous papers (1)—referred to as Mass I. and Mass II.— the question of the definition and observational determination of mass in Newtonian mechanics was studied. It is proposed here to consider some further aspects of the problem.

Jeffreys (2) has developed the subject of mass by a method somewhat different from that followed by Mach (3) and Pearson (4). On the other hand, Filon (5), who formerly supported the views of Mach and Pearson, expressed in his presidential address to the Mathematical Association read in 1938 a few days after his death in December 1937, the opinion that Mach's argument regarding the ratio of the masses of two bodies was invalid. Lastly, in a paper evoked by Mass I. and published simultaneously with Mass II., Narlikar (6) has expressed some views on this question.

In this paper the modes of approach of these writers are discussed and their connexion with that of the present writer is considered.

It will be useful here to emphasize the angle from which the subject is approached in Mass I. and Mass II. The usual aim of a dynamical investigation is to calculate the motions of bodies with prescribed forces and masses, subject to prescribed initial conditions. The reverse question has been considered in Mass I. and Mass II. That is to say, the problem for consideration was: assuming that complete knowledge of the motion of a system of bodies (particles) is available for an interval of time, to what extent is it possible to determine their masses (or rather the ratios of their masses) and also the forces between them? This mode of approach is expressed by saying: "An observer cannot observe the ratios of the masses of particles or the reactions between them: he can only aspire to infer them."

The present paper consists of three parts. The writings of Jeffreys, Filon, and Narlikar are discussed in Part I., Part II., and Part III.

---

* Communicated by Professor W. M. Smart.

† The references and the abbreviations here used are given on p. 484.

respectively. The bearing of the result in Part I. of Mass II. on the idea of Mach is also briefly considered in Part II. below.

PART I.

§ 1. *Jeffreys's Method.*

Jeffreys has developed the subject of dynamics as a consequence of the principle that "it is possible to learn from experience and to make inferences from it beyond the data directly known by sensation." He commences the discussion with an examination of the modes of taking terrestrial and celestial observations that are possible. He then obtains by a process of induction second order differential equations incorporating the principle of the conservation of momentum for a binary star. Subsequently he considers the case of the solar system. By a process of successive approximations and by considering "quasi-isolated" systems he formulates the law of gravitation and evolves the notion of mass. He also considers the process of balance weighing. Ultimately he extends the notion of mass thus derived to bodies inaccessible to experiment and observation.

The points of difference between the method of Jeffreys and those of Mach and Pearson are in brief: (1) that they use the third law of motion as a postulate and employ it to define mass, while Jeffreys seeks to establish them as a consequence of his guiding principle; (2) that Jeffreys is more concerned with the material systems amenable to experiment and observation rather than with abstract systems of dynamics than were the latter; and (3) that Pearson's attitude towards "force" is more theoretical than that of Jeffreys.

Jeffreys's method is, in general terms, similar to that outlined in example (3) on p. 1020 of Mass I. He has not, however, considered the theoretical status of mass and force when the system under consideration is such that it is not possible to think of "quasi-isolated" systems of two bodies as a first approximation.

PART II.

§ 2. *Filon's Experiments.*

Filon, who, in his first paper, supported Mach's concept of mass, came by about 1933 to the conclusion that the argument was fallacious. The reason which led to the investigation in his second paper was that "just as we can never know when a single body is under no force, so we can never know when two bodies are moving solely under their mutual reactions": in the basic form of Mach's definition of the ratio of the masses of two bodies it has to be assumed that they form an "isolated system."

On the question of definitions Filon has said: "It will probably be admitted that no definition of a physical quantity can be accepted as adequate unless it implies a criterion leading to an unambiguous measure of the thing defined." Before examining his argument, it will be useful to narrate three experiments which he has described. These can be conveniently called (*a*) the Fletcher-trolley experiment, (*b*) the falling pair experiment, and (*c*) the one body experiment. (*a*) was described in his first paper, (*b*) and (*c*) being in the second. His conclusions from (*c*) are also stated.

(a) *The Fletcher-trolley Experiment.*

Two bodies A_1 and B_1, which are connected by an elastic band, can move in a horizontal plane. After the motion is started, their simultaneous accelerations at a number of instants can be determined by means of an ingenious recording device. It is found that the ratio of their accelerations is constant.

In so far as the experiment is concerned, A_1 and B_1 move solely under the action of their mutual reactions. Hence, Mach's concept can be applied directly.

(b) *The Falling Pair Experiment.*

Two bodies A_2 and B_2 are connected by a light spring and are held in a horizontal straight line. They are then released and allowed to fall freely under the action of gravity and the spring. The line A_2B_2 remains horizontal during the motion. By noting the horizontal distances traversed by the bodies, it is possible to find out the horizontal components of their accelerations; their negative inverse ratio, found to be constant, would be the ratio of their masses. For, when the bodies remain in a horizontal straight line, the horizontal components are the mutually induced accelerations envisaged by Mach; the forces on each body are the tension of the spring, which is horizontal, being the reaction between the two bodies, and weight, which acts vertically downwards.

Filon called (*b*) "Mach's experiment"; this term can be applied to (*a*) also. He classed (*b*) as "comparatively simple in theory, though unsuitable in practice." In (*b*) the bodies A_2 and B_2 do not form an "isolated" system. Consequently, Mach's concept cannot be applied to it in the basic form. It can be employed in the case when the bodies do not remain in a horizontal straight line, only because the acceleration due to the earth's field is known.

(c) *One Body Experiment.*

It is assumed, to start with, that a body falls freely *in vacuo* at the surface with a constant acceleration g directed along the downwards vertical, and that

$$\vec{W}=m\vec{g} \quad \ldots\ldots\ldots\ldots \quad (1)$$

where $\vec{W}$ is the field force due to the earth, the mass m being undefined.

A body is suspended by a light string from a fixed support and the tension T_1, which balances W acting vertically downwards, is noted. This value of W, and consequently that of T_1, is termed the *rest weight* of the body and is denoted by w. It has been pointed out that w is not necessarily equal to mg.

The point of support is then made to move vertically with a vertical acceleration f, measured positive downwards, and the tension of the string is noted. He has described an ingenious arrangement for measuring T and controlling the acceleration f.

When the experiment is carried out for different values of f for the same body, it is found that T_1-T is equal to λf, where λ is constant and equal to w/g, g being known.

Consequently,

$$W=T+mf=T_1+(m-\lambda)f=w+(m-\lambda)f. \quad \ldots\ldots \quad (2)$$

The relation $\lambda=w/g$ follows from eqn. (2) by putting $f=g$ in it and using eqn. (1). Hence, the relation for vertical motion becomes

$$W=w+(m-w/g)f. \quad \ldots\ldots\ldots \quad (3)$$

From further experiments with an inclined plane and pulley, he arrives at the conclusion that in all cases it is necessary to assume that the requisite field force is obtained by adding to the rest weight a force $(m-w/g)f$ in the same direction as the acceleration f, which may have any direction.

After this stage he proceeds as follows. Let $\vec{F}$ be the contact force on the body; then

$$m\vec{f}=\vec{F}+\vec{W}=\vec{F}+\vec{w}+(m-w/g)\vec{f},$$

i. e.,

$$\frac{w}{g}\vec{f}=\vec{F}+\vec{w}. \quad \ldots\ldots\ldots\ldots \quad (4)$$

From eqn. (4) he concludes that everything happens as if the field force were equal to w, the rest weight, with w/g written instead of mass. He points out that the inertial mass m has disappeared from eqn. (4) and that it is undetermined and apparently undeterminable. He goes on to remark that Mach's experiment can only give the ratio of the rest weights. To sum up, he writes : " Mach's procedure for determining the ratio of the inertial masses without introducing gravity is illusory, and we have, in fact, been brought back by a devious route to Newton's definition, which appears to be the only possible one."

§ 3. *An Analysis of Filon's Investigation and Conclusion.*

The aspect on which Filon based his objection to Mach's concept, namely, the impossibility of knowing when two bodies form an "isolated" system, was cited in § 2 above. This question was considered in Part I. of Mass II.; the criterion based on the laws of motion was formulated and its consequences were worked out in relation to Mach's ratio.

There is also a possible objection to Mach's concept that it is generally not possible, leaving out the case of double stars, to have a system of only two bodies forming an "isolated" system. The answer to such an objection is that it is often necessary to envisage ideal systems and conditions, which may not actually obtain in the prescribed and imagined form, but which may be approached in some cases to an extent sufficient to justify the utilization of the abstract notions. Their usefulness consists in that they give a coherent and logical basis for a concept and that they can be used for a logical extension to more complicated cases. The position of Mach's idea in relation to a system of more than two particles was examined in § 3 of Mass I.; and it was shown that it was possible to say that the ratios of the masses of all pairs of particles in the system were *conceptually* perfectly definite when Mach's original idea was employed. In the light of the result in Part I. of Mass II. it became necessary to qualify the foregoing statement by saying that the ratios were conceptually perfectly definite *relative* to the Newtonian frame of reference in question.

The principal feature of Filon's remarkable investigation is that it gives a theoretical correlation of weight and gravity with the equations of motion. It is essentially concerned with terrestrial mechanics and introduces a new axiom that the "spring balance" measure of force is invariable and absolute. His method is not serviceable from the point of view of Mass I. and Mass II., since forces are not observable, and since subsidiary experiments on a system of particles like that of the spring balance cannot be performed.

Towards the close of his second paper he said: "I cannot to-day enter into the question of celestial observations, except to say that careful analysis of their implications does not apparently disturb our conclusions." It is rather difficult to anticipate exactly which analysis Filon had in his mind. This question has, however, been discussed in Mass I. and Mass II. from the point of view adopted there.

As regards Filon's view that, in the light of his analysis, Mach's ratio would mean the ratio of the rest weights of bodies, it is doubtful if this would necessarily impair that definition. It is quite possible to assert that once the ratio defined by the mutually induced accelerations, which remains constant in time, is obtained, it can be taken to be the ratio of the inertial masses, because this is what it is from the third law of motion, and that the connexion with rest weights comes in only because the investigation under consideration is confined to terrestrial mechanics *. It appears, therefore, that Filon's analysis neither proves nor disproves Mach's idea.

The important feature of Mach's concept † is that it gives a *conceptual*, but dynamical or rather kinematic, meaning to mass in the theoretical structure of the mechanics of any system of particles (bodies) in so far as the frame of reference is concerned. A possible source of danger to his idea lies, however, in the result established in Part I. of Mass II., because the ratio as defined by Mach does not have the same value for all observers. This would appear to mean that if it be desired to have a standard specification, so to speak, of the ratio of the masses of two or more than two particles, it would be necessary to single out one observer and the set of observers moving with uniform velocity relative to the former as the fundamental set of observers and thus to invoke something very similar to absolute space. This is a disconcerting conclusion. It is to be hoped that it may be possible to dispense with it.

PART III.

§ 4. *A Preliminary Explanation concerning Mass I. in Relation to Narlikar's Paper.*

In his paper evoked by Mass I. and published simultaneously with Mass II., Narlikar stated that the present writer pointed out in Mass I. two lacunae in the definition and determination of mass. The lacunae in the order followed by Narlikar are

(i.) That it is necessary to postulate that the mass ratios of bodies, as determined by their mutual interactions, are constant ; and

---

* In the one body experiment Filon studied the theoretical scheme for the motion of *one* body in the field of the earth and under the action of contact forces ; he extended the conclusions drawn in this manner to *two* bodies. Consequently, it is to be expected that the rest weight would make its appearance in the value of the kinematic Mach ratio. Throughout the investigation he treated mass as an undefined number which disappeared from the equation of motion for one body in its final form and was effectively replaced by w/g. This investigation included gravity and weight everywhere. When Filon's procedure is compared with that of Mach, it is seen that the latter *started* with the question of two bodies forming an " isolated " system and then applied to it the third law for the derivation of the definition of the ratio of masses of two particles.

† When Mach spoke of " mutually induced accelerations," he probably had in his mind also the case of gravitation in a generalized sense, *i. e.*, wider than astronomical.

(ii.) that Mach's idea does not enable an observer to compute the ratios of the masses when there are more than a certain number of bodies (particles) in the system.

Narlikar proposed a method of removing them. Before considering his views, it would be useful to give some explanation.

The postulate forming the first lacuna, stated above, was not represented as a lacuna. It was mentioned in § 2 of Mass I. in order that there should be a full, though necessarily brief, narrative of Mach's work. In Mass I., however, the relation of the Mach ratio to the frames of reference, *i. e.* observers, was not discussed. In Part I. of Mass II. it is shown that if there exists one observer who can say that a pair of particles forms an "isolated" system and can thus determine the ratio of their masses, following Mach's idea, which is the only weapon at his command, there exists an infinity of observers each of whom can assert that the pair is an "isolated" system and that the ratios determined by different observers differ from one another. It is not possible to say that any one value is more fundamental than any other. This effectively means that Mach's idea does not lead to a mass ratio invariant for all observers.

§ 5. *An Analysis of Narlikar's Suggestions* *.

Narlikar introduces the equations

$$(\mathrm{A})\,\frac{du}{dt}=0, \quad (\mathrm{B})\,\frac{dm}{dt}=0, \qquad \ldots\ldots \quad (1)$$

instead of the single equation

$$\frac{d}{dt}(mu)=0. \qquad \ldots\ldots \quad (2)$$

The equations refer to the motion of *one* particle under no force and any two of (A), (B) and (2) imply the remaining one. (B) simply posits the constancy of *m*. Moreover, it appears as if he is referring to absolute space of some sort.

Passing at this stage to the theory of relativity †, he points out that (2) corresponds to

$$\frac{d}{ds}\left\{m\frac{dx^\mu}{ds}\right\}+m\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds}=0, \qquad \ldots\ldots \quad (3)$$

which implies that

$$(\mathrm{A}')\,\frac{d^2x^\mu}{ds^2}+\Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{ds}\frac{dx^\beta}{ds}=0, \quad (\mathrm{B}')\,\frac{dm}{ds}=0. \qquad \ldots\ldots \quad (4)$$

---

* The numbering of equations in this section is similar to that of Narlikar.

† (A) and (B) above are the analogues of (A′) and (B′) respectively below.

It is to be pointed out that this only implies that the invariant mass of *an isolated* particle remains constant (7). It does not appear permissible to assert on the basis of the result for *one* isolated particle that the Mach ratio for *two* particles remains constant* , as Narlikar seems to have done.

It may be added that both Mass I. and Mass II. referred to Newtonian mechanics.

At the end of his paper, Narlikar has indicated how the masses of the particles of a system can be calculated with the help of the inverse square law. It was indicated in Mass I. how the law of force between the particles of an " isolated " system and the ratios of their masses can be determined by a process of approximation and subsequent induction : the process is applicable in certain circumstances only. The solar system was also specifically considered in Mass I. Furthermore, a detailed discussion of the computation of masses was given in § 4 of Mass II. ; and the possibility of the computation of masses, when a law of force has been assigned, has not been disputed.

In conclusion, it has to be emphasized that the object of Mass I. and Mass II. has been to examine the theoretical status of mass and, incidentally, force in relation to observers and observations, and not mere computation of prescribed constants for one particular type of motions.

References.

(1) C. G. Pendse :
  (i.) " A Note on the Definition and Determination of Mass in Newtonian Mechanics," Phil. Mag. (7) xxiv. pp. 1012–1022 (1937).
  (ii.) " A further Note on the Definition and Determination of Mass in Newtonian Mechanics," Phil. Mag. (7) xxvii. pp. 51–61 (1939).
    (i.) and (ii.) are referred to as " Mass I " and " Mass II " respectively.
(2) H. Jeffreys, ' Scientific Inference ' (1931). Preface and chapter viii.
(3) E. Mach, ' The Science of Mechanics,' ed. 4 (1919).
(4) K. Pearson, ' The Grammar of Science,' ed. 2 (1900), chapter viii.
(5) L. N. G. Filon :
  (i.) " Some Points on the Teaching of Rational Mechanics," ' The Mathematical Gazette,' xiii. pp. 146–153 (1926).
  (ii.) " Mass and Force in Newtonian Mechanics," ' The Mathematical Gazette,' xxii. pp. 9–16 (1938).
    (i.) and (ii.) are referred to as the first paper and the second paper respectively (of Filon).
(6) V. V. Narlikar, " The Concept and Determination of Mass in Newtonian Mechanics," Phil. Mag. (7) xxvii. pp. 33–36 (1939).
(7) A. S. Eddington, ' The Mathematical Theory of Relativity,' ed. 2, p. 127 (1924).

c/o Downing College,
Cambridge.

---

* *Cf.* also Part I. of Mass II.

XLVIII. *Iron-cored Coils of High Efficiency.*

By G. F. Partridge, B.Sc., F.Inst.P.*

[Received March 22, 1939.]

Inductances of relatively low resistance are of particular importance in the construction of filter circuits and tuned circuits generally. As is well known, the lower the power factor of a coil the more efficient it will be when used in a tuned circuit. For this reason it has been suggested that the reciprocal of the power factor is the most suitable criterion of the efficiency of an inductance when used for this purpose(1). This has been termed the magnification factor and is represented symbolically as $L\omega/R$, where L and R are the inductance and effective resistance measured at a pulsatance ω.

At wireless frequencies it is possible with good design to obtain air-cored coils of reasonable size with $L\omega/R$ values of the order of 200. At acoustic frequencies air-cored coils of such an efficiency become unduly large, and where economy of space is desirable iron-cored coils can be used with advantage.

The efficiences of such coils at the lower audio frequencies is the subject of this paper.

Continuous Iron Core.

We will first consider the efficiency of coils designed for a definite frequency, in which case we will omit the frequency term and find an expression for the time constant L/R.

The effective resistance of an iron-cored coil is the sum of two resistances, R_e due to the copper losses in the winding and R_i due to the losses produced by hysteresis and eddy current effects in the iron.

As regards the iron losses it can be assumed that, at low flux densities, the total losses are very nearly proportional to the square of the flux density(2).

Thus we may put

$$W=\kappa . B^2 . v,$$

where W=iron losses in watts,

B=flux density in gauss,

---

* Communicated by the Author.

κ=constant for the particular iron,

and v=volume of iron in c.c.

If R_i is the equivalent series resistance representing iron losses$^{(2)}$ and i is the current in the coil, we have

$$R_i=\kappa\left(\frac{B}{i}\right)^2 v \text{ ohms.}$$

Also

$$L=\frac{0\cdot4\pi N^2\mu A\times10^{-8}}{l_i} \text{ henries,}$$

$$B=\frac{0\cdot4\pi N\mu i}{l_i},$$

where N=number of turns in the winding,

A=cross-sectional area of the iron core in cm.^{2},

l_i=mean length of iron path in cm.,

and μ=permeability of the iron,

so that

$$R_i=\frac{\kappa(0\cdot4\pi)^2N^2\mu^2A}{l_i} \text{ ohms,}$$

and

$$\frac{L}{R_i}=\frac{10^{-8}}{0\cdot4\pi\kappa\mu} \quad \ldots\ldots\ldots\ldots \quad (1)$$

$$=\frac{8\times10^{-9}}{\kappa\mu} \text{ very nearly.}$$

This ratio is independent of the number of turns and also of the dimensions and inversely proportional to κ and μ. In order therefore to obtain a coil with large time constant both κ and μ must be small. In other words, the iron of which the core is composed should have low losses per unit volume together with low permeability.

In recent years various alloys have been introduced for sheet magnetic materials. Thus we have the silicon steels and the group of nickel iron alloys known as permalloys, all of which have much lower energy losses than magnetic iron. In Table I. are given a few representative

TABLE I.

Material.	Iron losses watts per c.c.	μ.	κ.	$\frac{L}{R_i}=\frac{8\times10^{-9}}{\mu\kappa}$.
Silicon steel	0·013	2500	$32{\cdot}00\times10^{-10}$	$1{\cdot}0\times10^{-3}$
Permalloy "B"..	0·0037	10000	$8{\cdot}15\times10^{-10}$	$0{\cdot}98\times10^{-3}$
Permalloy "C"..	0·0014	75000	$3{\cdot}10\times10^{-10}$	$0{\cdot}35\times10^{-3}$

figures for sheet material 0·015 inch thickness at B=2000 gauss, and a frequency of 200 c.p.s.

From the figures it will be seen that the relatively low energy losses obtained with permalloy are associated with higher permeability, so that the L/R_i obtained with lower loss material is actually smaller.

The values obtained for L/R_i can be taken as representing the order of values of L/R, since it can be shown that in general R_c is negligible compared with R_i. It is apparent, therefore, that it is impossible to obtain an efficient coil with a *continuous* iron core; in fact, in some cases the coil with the iron core removed will be the more efficient.

Iron Core with an Air-gap.

If there is an air-gap in the iron circuit the expression for L/R_i becomes

$$\frac{L}{R_i}=\frac{10^{-8}}{0{\cdot}4\pi\kappa}\left[\frac{1}{\mu}+\alpha\right], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where
$$\alpha=\frac{\text{length of air-gap}}{\text{length of iron path}}.$$

Now with low loss materials $1/\mu$ is negligible compared with α except for very small gaps, so that we have

$$\frac{L}{R_i}=\frac{8\times10^{-9}\alpha}{\kappa}\text{ very nearly.} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

TABLE II.

Material.	κ.	$\frac{L}{R_i}$.
Silicon Steel	$32{\cdot}00\times10^{-10}$	0·0125
Permalloy "B"	$8{\cdot}15\times10^{-10}$	0·049
Permalloy "C"	$3{\cdot}10\times10^{-10}$	0·129

Since this expression is independent of μ it is obvious that with low values of κ, *i. e.*, with cores of low loss material the introduction of an air-gap leads to improved values of L/R_i. This can be seen from Table II., in which a few values have been given for $\alpha=0{\cdot}005$.

The complete expression for the time constant must include the resistance R_c due to copper losses. This for a given shape of winding is proportional to the square of the number of turns and inversely proportional to the dimensions, so that we may put

$$R_c=K.\frac{N^2}{l_i},\text{ where K is a constant,}$$
$$=k_1N^2;$$

also

$$L=\frac{0\cdot4\pi N^2A\times10^{-8}}{l_i/\mu+l_a}$$

$$=k.\frac{N^2}{l_a},$$

since $\frac{l_i}{\mu}$ is small compared with l_a, the length of the air-gap.

$$B=\frac{0\cdot4\pi Ni}{l_a},$$

$$R_i=\frac{\kappa(0\cdot4\pi)^2N^2v}{l_a^2}$$

$$=k_2\frac{N^2}{l_a^2},$$

and

$$\frac{L}{R}=\frac{L}{R_c+R_i},$$

$$=\frac{k/l_a}{k_1+k_2/l_a^2}, \qquad (4)$$

where k, k_1, and k_2 are constant for a given coil.

From the form of this expression it will be apparent that as l_a is varied L/R passes through a maximum value for a certain "optimum" length of gap, and by differentiation this can be shown to be given by the following

$$l_a\,(\text{optimum})=\sqrt{\frac{k_2}{k_1}}. \qquad (5)$$

The fact that the time constant of an iron-cored coil, having an air-gap, can be improved by adjustment of the gap under the conditions stated above is very interesting.

This result is not to be confused with the "optimum gap effect" in an iron-cored coil carrying alternating current superposed on direct current (3) and (4). This is well known and is frequently used in designing choke coils. In this case the marked improvement in inductance produced by using a suitable air-gap is due to the effect of the air-gap upon the incremental permeability.

If in the present case there were direct current also passing through the coil it would be necessary to consider both effects.

As illustrating the above results curves are given in fig. 1 showing the variation of L/R and of R_i with length of gap for an iron-cored coil, the volumes of the iron and the winding being 27·7 c.c. and 20·2 c.c. respectively.

The maximum value of L/R which can be obtained by adjusting the gap is found by inserting the value of l_a (optimum) in the expression for L/R.

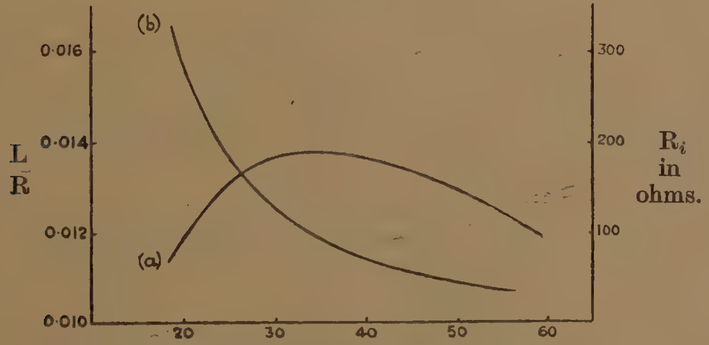
We have from (4) and (5)

$$\frac{\mathrm{L}}{\mathrm{R}}(\text{max.}) = \frac{k}{2\sqrt{k_1k_2}}$$

$$= \frac{10^{-8}\sqrt{\mathrm{A}}}{2\sqrt{\mathrm{K}.\kappa}}. \quad \ldots\ldots \quad (6)$$

In the above expression K is constant for a given shape of coil and κ is constant for a given core material. For a given shape of coil therefore L/R is proportional to $\sqrt{\mathrm{A}}$, *i. e.*, for a symmetrical change in the size of the coil, the value of L/R varies directly as the linear dimensions. This assumes that the space factor for the insulation of the winding is the same for the gauges of wire used. If small gauges of wire are involved then, in general, the L/R value will increase more rapidly than the dimensions due to the improvement in space factor with the thicker wire.

Fig. 1.



Length of gap in cm. × 10^{-3}.

If the size of the coil is fixed, then to obtain the best time constant it is necessary to select the shape which gives the greatest value for the ratio A/K, since under these conditions K is not constant.

Many advantages are claimed for the toroidal form of winding. Where very good magnetic screening is required this form of winding is not very suitable for the type of coil suggested. For it would be impossible to adjust the air-gap with a toroidal winding if fully wound.

It will appear from figures given later that coils with built-up cores of the usual types of transformer stampings can be designed for frequencies at least up to 2000 c.p.s., which have time constants comparable with those obtained with the toroidal form of winding.

Variation with Frequency.

In considering the efficiency of iron-cored coils over a range of frequencies it will be assumed that R_c is constant and that up to frequencies of about 2000 c.p.s. the iron losses are chiefly hysteresis losses, so that κ is approximately proportional to frequency.

The expression for $L\omega/R$ will then be of the following form:

$$\frac{L\omega}{R}=\frac{Sf/l_a}{S_1+S_2f/l_a^2}$$

where S, S_1, and S_2 are constants for a given shape of coil.

The variation of $L\omega/R$ with frequency will thus depend upon the frequency for which l_a has been adjusted to give the maximum value of L/R.

If this frequency is f_0, then

$$l_a^2=\frac{S_2}{S_1}\cdot f_0$$

and

$$\frac{L\omega}{R}=\frac{Sf\sqrt{f_0}}{\sqrt{S_1S_2}(f_0+f)}. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (7)$$

In order to demonstrate the effects of gap upon the variation of $L\omega/R$ with frequency, theoretical values have been determined from the expression (7) for a coil occupying a space of approximately 100 c.c. The volumes of iron and of the winding involved are 22·5 c.c. and 40 c.c. respectively. The iron is assumed to be permalloy "C."

For the particular shape chosen the constants have the following values:

$$S\ =1{\cdot}18\times10^{-6},$$

$$S_1=5{\cdot}58\times10^{-7},$$

$$S_2=6{\cdot}0\ \ \times10^{-11}.$$

The results are shown graphically in fig. 2 for four values of f_0, and the corresponding values of l_a and α, the air-gap ratio, are given in Table III.

In order to test the validity of these results measurements have been made of the efficiency of a coil designed on these principles. The volume of the iron in this coil is 22 c.c. and that of the winding is 52 c.c.

The results of the measurements are given in fig. 3. It will be noticed that the type of variation obtained is very similar in character to that given by theory. The values of $L\omega/R$ obtained are not quite as large as the theoretical for the same volume of coil. This is partly due to the fact that the available stampings were not quite suitable for the best design. Also there is evidence of a slight departure from linearity in the variation of R_i with frequency, as shown in fig. 4. Thin stampings were used so that the eddy current losses were relatively small.

Composition Cores.

Within recent years various types of composition cores such as Ferrocart and Gecalloy have been introduced, particularly for use at high frequencies. These materials have the property of low losses together with low effective permeability. They consist essentially of small particles of magnetic material embedded in a non-magnetic matrix.

Fig. 2.

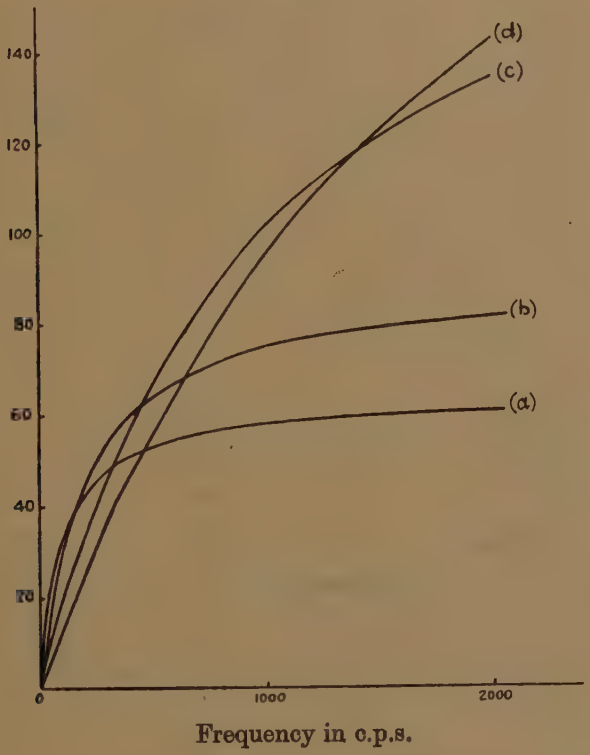


TABLE III.

Curve.	f_0 in c.p.s.	l_a (in cm.).	a.
(a)	100	0·033	0·0022
(b)	200	0·046	0·0031
(c)	1000	0·104	0·0069
(d)	2000	0·147	0·0098

Fig. 3.

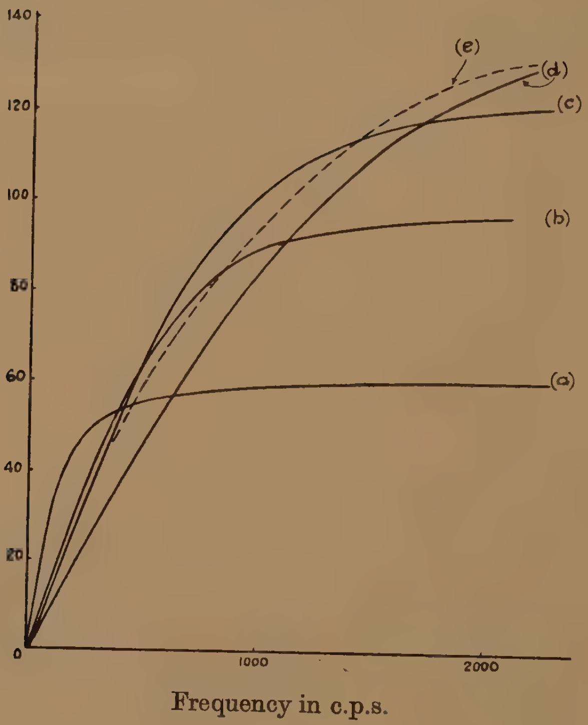
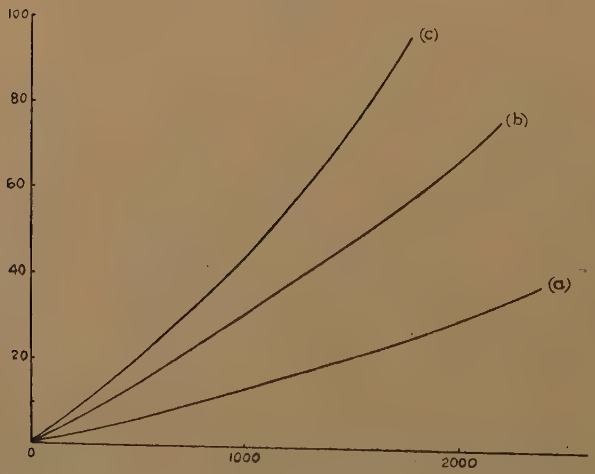


Fig. 4.



(a) f_0=2000 c.p.s. (b) f_0=950 c.p.s. (c) f_0=160 c.p.s.

It is interesting to note that the low effective permeability is due to the separation of the particles, and that this is equivalent in principle to the introduction of an air-gap as described above. In other words, the material behaves exactly as if the magnetic particles were separated by air-spaces.

If we assume that the particles are of uniform size and symmetrically distributed it is possible to obtain an expression for the effective permeability of the material.

We will assume that in any linear direction the proportion of air-space to magnetic material is α, so that α has the same significance as before. The permeance of any path of length l and cross-sectional area A is then given by

$$\text{permeance}=\frac{1}{\mathrm{P}}=\frac{\mathrm{A}}{l(1+\alpha)}\left[\frac{\mu}{1+\mu\alpha}+\frac{2\alpha+\alpha^2}{1+\alpha}\right],$$

so that the effective permeability of the material is *

$$\mu_{\text{eff}}=\frac{1}{1+\alpha}\left[\frac{\mu}{1+\mu\alpha}+\frac{2\alpha+\alpha^2}{1+\alpha}\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Similarly, or by comparing expressions (1) and (2), it can be seen that the effective permeability of an iron core with an air-gap is

$$\mu_{\text{eff.}}=\frac{\mu}{1+\mu\alpha}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

For small values of α, such as those given in Table III., the expression (8) for the effective permeability of a composite core reduces to that for an iron core with an air-gap. Using this expression and assuming that the magnetic material is permalloy C at low flux densities with a permeability of 20,000, we find that these values of α correspond to composition cores having effective permeabilities from about 450 to 100. Composite cores for high-frequency coils are usually of much lower effective permeability—of the order of 15 or less which corresponds to an air-gap ratio of about 0·07. This, together with the fact that the particles are of small size, accounts for the low iron losses.

At the frequencies considered here the eddy current losses in iron-cored coils can be made small by using thin sheets of iron. The efficiencies which can be obtained are for the same volume of coil equally as good as those obtained with the more expensive composite cores. As illustrating this a curve is given in fig. 3—see dotted curve (*e*)—for a well-known make of composite core toroidal coil, having an overall volume of 100 c.c.

---

* A similar type of expression has been given by Howe, 'Experimental Wireless,' May 1933.

As shown in the previous section, the frequency characteristic of an iron-cored coil can be adjusted, whereas that of a composite core is fixed by the distribution of the iron in the core and can only be adjusted by introducing a variable air-gap or by varying the composition of the core.

Summary.

The efficiency of laminated iron-cored coils as represented by their $L\omega/R$ value is considered. It is shown that, in order to take advantage of the low losses obtained with permalloy cores, it is necessary to use an air-gap. For any given frequency there is an " optimum gap " which gives maximum efficiency. Formulæ are given for calculating the value of this gap, and also the efficiency of any given coil. Following from this it is shown that the frequency characteristic of such a coil will depend upon the value of the gap, in that it will be determined by the frequency for which the gap has been adjusted for the optimum condition.

Theoretical curves are given and measurements confirming the results for frequencies up to 2000 p.p.s.

It is shown that the same principle accounts for the high efficiencies obtained by using composite cores such as Ferrocart and Gecalloy.

Acknowledgment.

The author is indebted to the Admiralty for permission to publish this paper.

References.

(1) S. Butterworth, " Design of Low Loss Coils," ' Experimental Wireless,' July 1926 ; and ' Wireless World,' December 1926.
(2) D. W. Dye, " Performance and Properties of Telephonic Frequency Intervalve Transformers," ' Experimental Wireless,' September 1924.
(3) C. R. Hanna, " Design of Reactances and Transformers which carry Direct Current," Journal American Inst. E.E., February 1927.
(4) G. F. Partridge, " Inductance of Iron-cored Coils carrying Direct Current," Phil. Mag., January 1937.

XLIX. *The Equations of Motion of Point Electrons deduced from a Variational Principle.*

By T. LEWIS, Aberystwyth *.

[Received January 8, 1940.]

ATTENTION has once again been focused on the laws of motion of a classical type of electron by the recent papers of Dirac (1938) and Pryce (1938). These authors regard the electron as a structureless point, a singularity of the electromagnetic field, the latter being described everywhere except at these singularities by Maxwell's equations for free aether. Dirac gets over the difficulty of the infinite electromagnetic energy and momentum associated with such a model by attributing to the electron an infinite negative mass concentrated at the electron itself, which neutralizes the infinite electromagnetic mass and leaves a little to spare. This surplus he identifies with the inertial mass of the electron. Pryce, on the other hand, associates with each electron an energy momentum field in addition to the ordinary electromagnetic energy and momentum. This new energy momentum field depends upon the charge, and when superposed on the electromagnetic energy-momentum field it cancels the more troublesome terms in the latter. One strange feature of this new field is that it contains no term which could represent the mutual energy of two electrons as is the case with electromagnetic energy. Pryce calls the sum of these two energies (momenta) the electromagnetic energy (momentum). In addition to this extra energy-momentum field Pryce also introduces a mechanical energy-momentum vector associated with the electron itself. Both Dirac and Pryce arrive at well-defined equations governing the motion of an electron. Their general argument, however, does not lead unambiguously to these equations.

The object of the present paper is to investigate the behaviour of point electrons in one another's presence, without introducing concepts which are foreign to classical electromagnetism. We make no use of the concept that energy and momentum are localized in the field, and it is not surprising, therefore, that our final equations of motion of an electron should differ from those based on this concept. The so-called reaction of radiation is missing from our equations. But this is a term which first came into the equations of motion as part of the force which an electron exerts on itself, the electron being regarded as finite in extent.

---

* Communicated by the Author.

It is due directly to retarded action. The whole method of calculation breaks down when the electron is a geometric point. Again, the localization of energy and momentum in the field was made plausible on account of the Lorentz principle of motion applied to a Lorentz type of electron. This principle has no meaning for point electrons. (Objections to our equations are dealt with in paragraph 5.)

It is well known that half of Maxwell's equations are satisfied by expressing the field intensities **E** and **H** in terms of the electromagnetic potentials (ϕ, **a**), and the other half are contained in the variational principle $\delta I=0$, where

$$I=\frac{1}{8\pi}\iiiint (H^2-E^2)\,dx\,dy\,dz\,dt.$$

Now the field depends upon the motion of electrons, or, to use the language of relativity, upon the world lines of electrons. If we assume one of these world lines to be slightly displaced it will cause a slight change in the field at all points, and hence, in general, a change in I. It therefore seems not unreasonable to expect that the actual "positions" of the world lines of electrons in space-time are such as to make I stationary for these positions as compared with infinitely near positions. But, as we shall see later, the displacement of a world line must be one which leaves the 'lengths' of all its elements unchanged, otherwise the change in I becomes infinite and hence meaningless.

We therefore postulate that the laws of motion of electrons are contained in the variational principle

$$\delta I=0,$$

subject to a supplementary condition

$$\delta(ds)=0$$

for each electron, where δI denotes the increment in I due to the infinitesimal displacement of the world line of an electron.

We assume that the field depends only on the electrons present and can be calculated when the world lines of these electrons are prescribed. Again, we assume that retarded potentials only are effective, though our final equations show a complete symmetry between retarded and advanced potentials, a result which was also obtained by the author (Lewis, 1935) in the case of a spherical, perfectly reflecting, electron. Equations of this form were also obtained by Fokker (1929), though he does not proceed from pure field theory.

2. By the action of the electromagnetic field we shall mean the integral

$$I=\frac{1}{8\pi}\iiiint (H^2-E^2)\,dx\,dy\,dz\,dt, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

taken throughout the region of space-time bounded by the forward "light cones" * with vertices at the points $(0, 0, 0 \pm \mathrm{T})$, where T ultimately tends to infinity, and narrow tubes surrounding the world lines of electrons. Each of these tubes will be assumed to have a spherical spatial cross-section of constant radius ϵ for each value of the proper time, in a co-ordinate system in which the electron is instantaneously at rest. ϵ is assumed to be smaller than any physically significant length.

E and **H** are defined in terms of the scalar and vector potentials $(\phi, \mathbf{a})$ by means of the equations

$$\mathbf{E} = -\mathrm{grad}\, \phi - \frac{\partial \mathbf{a}}{\partial t}, \quad \mathbf{H} = \mathrm{rot}\, \mathbf{a}. \qquad (2)$$

It will be convenient to use relativistic notation, which is too well known to need detailed explanation. The covariant components of the field quantities are defined in terms of the four-vector A^μ by means of the formula

$$\mathrm{F}^{\mu\nu} = \frac{\partial \mathrm{A}^\nu}{\partial x_\mu} - \frac{\partial \mathrm{A}^\mu}{\partial x_\nu}, \qquad (3)$$

and

$$\mathrm{I} = \frac{1}{16\pi} \iiiint \mathrm{F}_{\mu\nu} \mathrm{F}^{\mu\nu}\, dx_0\, dx_1\, dx_2\, dx_3, \qquad (4)$$

where the contravariant components are formed with the aid of the fundamental tensor $g^{00} = 1$, $g^{11} = g^{22} = g^{33} = -1$, $g^{\mu\nu} = 0$, $\mu \neq \nu$.

If A^μ is given an arbitrarily small increment $\delta \mathrm{A}^\mu$ which vanishes on the boundary of the region of integration, the conditions for the vanishing of $\delta \mathrm{I}$ are

$$\frac{\partial \mathrm{F}_{\mu\nu}}{\partial x_\nu} = 0, \quad \mu = 0, 1, 2, 3, \qquad (5)$$

which are the remaining four of Maxwell's equations. It is usual to impose upon the A^μ the additional condition

$$\frac{\partial \mathrm{A}_\mu}{\partial x_\mu} = 0. \qquad (6)$$

On account of (5), the expression (4) for the action can be put in the form

$$\mathrm{I} = \frac{1}{8\pi} \iiiint \frac{\partial}{\partial x_\mu} (\mathrm{A}^\nu \mathrm{F}_{\mu\nu})\, dx_0\, dx_1\, dx_2\, dx_3. \qquad (7)$$

---

* The region bounded by these two cones will include any specified region of space-time provided T is made large enough. Since the action is transformed into surface integrals one cannot ignore the contribution of surfaces at infinity without investigation. One could equally well take the planes $t = \pm \mathrm{T}$ $(\mathrm{T} \to \infty)$ as boundaries—the labour is greater, but the result is the same.

Partial integration enables us to transform the action into integrals over the various boundaries. In general there will be three types of integral: (1) integrals taken over the tubes surrounding the world lines of electrons; (2) integrals taken over the bounding light cones; (3) an integral taken over a sphere of radius $R(\to\infty)$, and with respect to the time between the limits $\mp T+R$. We shall denote the corresponding contributions to I by I_e, I_c and I_s respectively, so that

$$I=I_e+I_c+I_s. \qquad (8)$$

I_s will never vanish if any of the electrons undergo acceleration at any time, but it will be shown in the appendix that as $R\to\infty$ I_s can be integrated with respect to the time and hence transformed into integrals over sections of the bounding light cones, so that instead of (8) we may write

$$I=I_e+I_c'. \qquad (9)$$

We are not interested in the absolute value of I, but merely in its variation due to the infinitesimal displacement of a finite portion of the world line of an electron. It is assumed that the field can be calculated when the world lines of all electrons are given—it is given by the *retarded* point potentials only. If, therefore, we vary a finite portion of the world line of an electron I_c' is not affected, for its value is derived from the state of the electrons near those points where their world lines cut the bounding light cones. We thus get the important result

$$\delta I=\delta I_e, \qquad (10)$$

where δI is the increment in I due to the small displacement of the world line of an electron.

It is therefore necessary to calculate I_e, the contribution of the tubes surrounding the world line of electrons.

Integrating the right-hand side of (7) we find that

$$I_e=\Sigma\frac{1}{8\pi}\iiint \lambda^\mu A^\nu F_{\mu\nu}\,dS, \qquad (11)$$

where $\boldsymbol{\lambda}$ is a unit vector along the outward normal to the three-dimensional surface element dS of a tube. The summation is extended over all electrons.

In order to take advantage of Dirac's calculations it will be convenient to use his notation. The world line of an electron will be described by the equations

$$z_\mu=z_\mu(s), \qquad (12)$$

or, in vector notation,

$$\mathbf{z}=\mathbf{z}(s), \qquad (13)$$

simply.

Differentiation with respect to s is denoted by dots, thus

$$\frac{d\mathbf{z}}{ds} = \dot{\mathbf{z}} = \mathbf{v}, \qquad (14)$$

where $\mathbf{v}$ is the velocity vector.

The scalar product of two vectors is defined by

$$(\mathbf{A}, \mathbf{B}) = A_\mu B^\mu = A_0B_0 - A_1B_1 - A_2B_2 - A_3B_3. \qquad (15)$$

We have, therefore,

$$(\mathbf{v}, \mathbf{v}) = 1,\ (\mathbf{v}, \dot{\mathbf{v}}) = 0, \text{ etc.} \qquad (16)$$

The general field point will be denoted by

$$\mathbf{x} = (x_\mu),\ \mu = 0, 1, 2, 3. \qquad (17)$$

The retarded potentials of an electron are given by

$$A_\mu = \frac{e\dot{z}_\mu}{(\dot{\mathbf{z}}, \mathbf{x}-\mathbf{z})}, \text{ or } \mathbf{A} = \frac{e\dot{\mathbf{z}}}{(\dot{\mathbf{z}}, \mathbf{x}-\mathbf{z})}, \qquad (18)$$

where the retarded proper time s is defined by

$$(\mathbf{x}-\mathbf{z}, \mathbf{x}-\mathbf{z}) = 0, \text{ with } x_0 > z_0. \qquad (19)$$

For a fixed $\mathbf{z}$, (19) is the equation of the light cone with vertex at $\mathbf{z}$.

The corresponding field quantities are given by

$$F_{\mu\nu} = -\frac{e}{(\dot{\mathbf{z}}, \mathbf{x}-\mathbf{z})}\frac{d}{ds}\frac{\dot{z}_\mu(x_\nu - z_\nu) - \dot{z}_\nu(x_\mu - z_\mu)}{(\dot{\mathbf{z}}, \mathbf{x}-\mathbf{z})}. \qquad (20)$$

In calculating the contribution of a single electron tube to I_e it is necessary to separate the field of that electron from the field of the remainder. Capital letters will refer to the electron's own field and corresponding small letters to the field of the remainder. Neglecting terms which vanish with ϵ, we find that the contribution of a single tube is

$$I_n = \frac{1}{2}\int\left\{-\frac{e_n^2}{\epsilon} - e_n(\mathbf{a}_n, \mathbf{v}_n)\right\}ds_n. \qquad (21)$$

The details of the calculation are given in the appendix so as not to interfere with the main argument.

The first term is due to the electron's own field and becomes infinite as ϵ becomes zero, and

$$\mathbf{a}_n = \sum_{m \neq n} \frac{e_m \dot{\mathbf{z}}_m}{(\dot{\mathbf{z}}_m, \mathbf{z}_n - \mathbf{z}_m)}, \qquad (22)$$

so that

$$I_e = \Sigma I_n = -\frac{1}{2}\sum_n \int \frac{e_n^2}{\epsilon}\, ds_n - \frac{1}{2}\sum_n \sum_{m \neq n} \int \frac{e_n e_m (\dot{\mathbf{z}}_m, \dot{\mathbf{z}}_n)}{(\dot{\mathbf{z}}_m, \mathbf{z}_n - \mathbf{z}_m)}\, ds_n. \qquad (23)$$

where in the last term

$$\dot{\mathbf{z}}_m = \frac{d\mathbf{z}_m}{ds_m}, \quad \dot{\mathbf{z}}_n = \frac{d\mathbf{z}_n}{ds_n},$$

and the proper times s_n and s_m are connected by the relation

$$(\mathbf{z}_n - \mathbf{z}_m, \mathbf{z}_n - \mathbf{z}_m) = 0, \text{ with } z_{n0} < z_{m0}. \quad \ldots \ldots \quad (24)$$

3. Let us now consider the increment in I_e due to the infinitesimal displacement of a portion of the world line of the nth electron which is consistent with the conservation of the "length" of all its elements, *i. e.*, subject to

$$\delta(ds_n) = 0. \quad \ldots \ldots \ldots \quad (25)$$

This condition is not arbitrary, but is a necessary one if the variation of I_e is to have meaning. From (23) we see that the first term of I_e becomes infinite and, in general, its first variation becomes infinite as $\epsilon \to 0$ unless the condition (25) is fulfilled. The effect of this condition is to leave the self-action of electrons unchanged. It follows, therefore, that we may ignore these terms entirely. But we must take into account (25) when we write down the equation

$$\delta I_e = 0. \quad \ldots \ldots \ldots \quad (26)$$

This is done in the usual way by multiplying (25) by an arbitrary multiplier k_n, integrating along the world line and adding to (26).

This gives rise to the equation

$$\int k_n \delta(ds_n) + \frac{1}{2} \sum_{m \neq n} \int \delta \frac{e_n e_m (\dot{\mathbf{z}}_m, \dot{\mathbf{z}}_n)}{(\dot{\mathbf{z}}_m, \mathbf{z}_n - \mathbf{z}_m)} ds_n = 0, \quad \ldots \ldots \quad (27)$$

which we postulate as the law governing the motion of the electron.

The last term can be put under a more convenient form. Denoting by I_{mn} the individual terms of the summation, we see that the terms involving the nth and mth electrons are $I_{mn} + I_{nm}$. These two terms are affected in different ways when the world line of the nth electron only is varied. I_{mn} involves the retarded potentials of the mth electron on the world line of the nth electron, while I_{nm} involves the retarded potentials of the nth electron on the world line of the mth electron. In fact,

$$I_{nm} = \frac{1}{2} \int \frac{e_m e_n (\dot{\mathbf{z}}_n, \dot{\mathbf{z}}_m)}{(\dot{\mathbf{z}}_n, \mathbf{z}_m - \mathbf{z}_n)} ds_m = \frac{1}{2} \int \frac{e_m e_n (d\mathbf{z}_n, d\mathbf{z}_m)}{(d\mathbf{z}_n, \mathbf{z}_m - \mathbf{z}_n)}. \quad \ldots \ldots \quad (28)$$

From (24) we obtain

$$(d\mathbf{z}_n, \mathbf{z}_n - \mathbf{z}_m) = (d\mathbf{z}_m, \mathbf{z}_n - \mathbf{z}_m). \quad \ldots \ldots \quad (29)$$

We thus see that

$$\mathrm{I}_{nm}=\frac{1}{2}\int\frac{e_m e_n(d\mathbf{z}_m,\, d\mathbf{z}_n)}{(d\mathbf{z}_m,\, \mathbf{z}_m-\mathbf{z}_n)}=\frac{1}{2}\int\frac{e_m e_n(\dot{\mathbf{z}}_n,\, \dot{\mathbf{z}}_m)}{(\dot{\mathbf{z}}_m,\, \mathbf{z}_m-\mathbf{z}_n)}\,ds_n, \quad . \quad . \quad . \quad (30)$$

where, since in this case $z_{m0}>z_{n0}$, the last form is to be interpreted as the integral of the advanced potentials of the mth electron along the world line of the nth electron.

Equation (27) may now be re-written :

$$\int k_n\delta(ds_n)+\frac{1}{2}\sum_{m\neq n}\int\delta\left[\frac{e_n e_m(\dot{\mathbf{z}}_m,\, \dot{\mathbf{z}}_n)}{(\dot{\mathbf{z}}_m,\, \mathbf{z}_n-\mathbf{z}_m)}\,ds_n\right]_{\text{ret.}}$$

$$+\frac{1}{2}\sum_{m\neq n}\int\delta\left[\frac{e_n e_m(\dot{\mathbf{z}}_m,\, \dot{\mathbf{z}}_n)}{(\dot{\mathbf{z}}_m,\, \mathbf{z}_m-\mathbf{z}_n)}\,ds_n\right]_{\text{adv.}}=0. \quad . \quad (31)$$

There is, of course, a corresponding equation for each electron.

When the variation is carried out we get four equations which should give some information about the arbitrary multiplier k_n. The process is too well known to justify giving it in detail here, so we merely write down the final equations for the nth electron, omitting the suffix n to avoid confusion. They are

$$\frac{d}{ds}(kv_\mu)=\tfrac{1}{2}ev_\nu(f^\nu_{\mu\,\text{ret.}}+f^\nu_{\mu\,\text{adv.}}),\quad \mu=0,\,1,\,2,\,3, \quad . \quad . \quad . \quad (32)$$

where f^ν_μ refers to the field of all the remaining electrons.

Multiplying by v_μ and summing over all values of μ we get

$$v^\mu\frac{d}{ds}(kv_\mu)=\frac{dk}{ds}=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

i. e., k is constant along the world line of the electron, and may be interpreted as its mass. Its magnitude thus far is indeterminate, but the equations of motion are given unambiguously in the form

$$k\dot{v}_\mu=\tfrac{1}{2}ev_\nu(f^\nu_{\mu\,\text{ret.}}+f^\nu_{\mu\,\text{adv.}}), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

which agrees with Dirac's equations when he makes the special assumption

$$\mathrm{F}_{\text{in}}+\mathrm{F}_{\text{out}}=0.$$

4. If now we consider two electrons, one with a mass which is very large in comparison with the mass of the other, a Lorentz system can be found in which the electron with the larger mass is comparatively at rest, so that its retarded and advanced potentials are more or less equal and fairly constant with respect to the time coordinate. The equations of motion of the other electron will then be approximately,

$$k\dot{v}_\mu=ev_\nu f^\nu_\mu. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

This is the relativity form of the equations governing the motion of an electron. The damping terms due to radiation are absent. This in no way proves that there has been an error in the derivation, for the missing terms are part of the force which an electron exerts on itself when it is assumed to occupy a small but finite volume. The same assumption allows us to regard energy and momentum as being localized in space and leads to the Poynting theory of radiation. We have made no appeal to these concepts, even were it legitimate to do so when the electron is regarded as a singularity of the field, and there is nothing to compel us to regard the Poynting vector as the rate of flow of energy and hence to introduce damping terms in (35). In fact, we claim equations (35) as the correct equations for a point electron in a theory which does not introduce concepts which are foreign to classical electromagnetic theory, such as the infinite negative mass of the electron itself introduced by Dirac to cancel the infinite positive mass of its field, or the supplementary energy-momentum tensor introduced by Pryce.

5. It has been urged by several critics that the absence of the radiation terms from the equations (35) renders their derivation valueless, though mathematically correct. This raises the question whether any investigation along classical lines has any value to-day, for it is generally admitted that classical electromagnetic theory cannot account satisfactorily for microscopic phenomena. A correct formulation of the classical equations is justified when, by applying the "recognized code," they can be translated into the corresponding correct quantum-mechanical equations. The classical Hamiltonian corresponding to (35) is precisely the Hamiltonian which has been used as the basis of the quantum theory of the electron in an electromagnetic field, and accounts for the main features of the line spectra of hydrogen-like atoms as well as a multiplicity of other phenomena.

One can imagine an objection raised against Dirac's and Pryce's equations on the ground that they do not contain terms representing the effect of the electron's spin, and therefore cannot give rise to the correct quantum-mechanical picture of the electron in a magnetic field. But in Dirac's own quantum theory of the electron, which is based on a classical picture not involving the spin, all the features of spin and magnetic moment are correctly accounted for without introducing special postulates as in the older quantum mechanics. This proves that the quantum-mechanical picture does not necessarily have a classical counterpart.

Again it is argued that the retardation due to damping is just what gives rise to the natural width of the spectral lines. This would be a valid argument if it could be shown that the Poynting theory of radiation can account for other features of the spectral lines, such as the frequency associated with them. When accounting for the natural width classical theory regards the electron as an oscillator. But it is known that the frequency of the emitted radiation is not the frequency of the oscillator, but is given by Bohr's relation $E' - E = h\nu$. In quantum mechanics the natural width of spectral lines follows from Heisenberg's "uncertainty principle" (Weisskopf, 1933). Our standpoint is not that emission of radiation by an electron would not modify its kinetic energy, but that classical radiation theory, in so far as it is embodied in the Poynting theorem, cannot give a satisfactory picture of radiation at all *.

The equations of motion of an electron containing radiation terms are not generally accepted. Frenkel (1926) states that on many grounds the equations (35) are to be preferred. Our equations for a system of electrons are identical with those derived by Fokker (1929). Scherrer (1939) uses the Lagrangian and Hamiltonian corresponding to our general equations of paragraph 3 as the basis of an investigation of dynamical models for heavy particles. The whole case is put very clearly by Milne, who has given me authority to quote him. He states :

" The classical formula for the rate of radiation from an accelerated electron is one of the least satisfactory formulæ in electromagnetism. It depends on the Poynting theorem, which depends on the localization of energy in the medium, which, in turn, depends on the continuous distribution of charge and is inapplicable to electrons considered as point singularities. Actually, there is no difficulty there at all, for once a charge is taken to be a singularity, the ordinary proof leading to energy $E^2/8\pi$ per unit volume is inapplicable. A point charge taken to be a singularity has no energy. Energy is only a mutual affair of two such charges in one another's presence.

" In the classical equation of motion of an electron in the presence of the field (E, H) there is no such term (radiation reaction). There is simply the Larmor-Lorentz force. If all the variations of E, H are taken into account, the reaction on the electron is automatically taken into account. It is not as though there were *both* the effect of the electromagnetic field on the electron *and* the reaction of the emitted radiation : whatever this " reaction," it is fully included if all the consequent changes in E, H are taken into account."

These arguments are given in greater detail in Milne's account of his own theory of electromagnetism (1938).

In this theory, which is based on point electrons, the field at a point has no meaning independent of the presence of an electron at that point.

---

* In absorption problems no appeal is made to the concept that field energy is flowing into the electron. One merely calculates the rate at which the electric force is performing work on the electron.

E and H are functions of the velocity of the electron. It is thus impossible to represent electromagnetic energy and momentum as space integrals, and Poynting's theorem is, therefore, meaningless. The energy equation shows that the rate of loss or gain of the energy of a system of electrons is linear in their accelerations. This is a very satisfactory feature of the theory, for it admits emission and absorption on the same footing.

To sum up, a classical investigation is justified if the results, on being translated by the "recognized code" gives rise to the correct quantum mechanical results. The classical picture need not of necessity contain, even qualitatively, all the counterparts of the quantum mechanical picture (the spin, for example). It is illogical to expect a classical theory to account for certain features of the line spectra when any classical theory is admittedly utterly inadequate to account for other, more important, features.

I wish to thank Professor E. A. Milne for the interest he has taken in this paper, and for his support of the point of view adopted in it.

APPENDIX.

In order to calculate the contribution of a single electron tube to I_e we follow Dirac and write

$$x_\mu = z_\mu(s_0) + \gamma_\mu, \qquad (36)$$

where γ_μ is small, in fact $\gamma_\mu \gamma^\mu = -\epsilon^2$. Derivatives will be taken at the proper time s_0. The retarded proper time at a point of the tube will be $s_0 - \sigma$ where σ is small, positive, of order ϵ. Considerable simplification results if $\boldsymbol{\gamma}$ satisfies the orthogonal relation

$$(\boldsymbol{\gamma}, \dot{\mathbf{z}}) = 0, \qquad (37)$$

for all values of s_0. Dirac (Eqn. (64), 1938) has shown that $\boldsymbol{\gamma}$ is normal to the surface of the tube. It follows at once that

$$\boldsymbol{\gamma} = \boldsymbol{\lambda}\epsilon. \qquad (38)$$

Now the term of highest order in A^μ involves ϵ^{-1}, and the term of highest order in $F_{\mu\nu}$ involves ϵ^{-2}. We shall disregard all terms which, when multiplied by ϵ^2 are of order ϵ, ϵ^2, etc., for they disappear when ϵ tends to zero. It follows that we need only retain the first two terms in both A^μ and $F_{\mu\nu}$.

We quote the following results from Dirac's paper.

$$\dot{z}_\mu(s_0 - \sigma) = v_\mu - \sigma \dot{v}_\mu + \dots \qquad (39)$$

$$(\dot{\mathbf{z}}, \mathbf{x} - \mathbf{z})^{-1} = \sigma^{-1}[1 - (\boldsymbol{\gamma}, \dot{\mathbf{v}})]^{-1}\{1 - \tfrac{1}{2}\sigma(\boldsymbol{\gamma}, \ddot{\mathbf{v}}) + \dots\} \qquad (40)$$

$$\sigma = \epsilon[1 - (\boldsymbol{\gamma}, \dot{\mathbf{v}})]^{-\frac{1}{2}}\{1 - \tfrac{1}{6}\epsilon(\boldsymbol{\gamma}, \ddot{\mathbf{v}}) + \dots\}. \qquad (41)$$

From these it is not difficult to obtain the expansion

$$\mathrm{A}_\mu=\frac{e\dot{z}_\mu}{(\dot{\mathbf{z}},\ \mathbf{x}-\mathbf{z})}=e\epsilon^{-1}[1-(\boldsymbol{\gamma},\ \dot{\mathbf{v}})]^{-\frac{1}{2}}\{v_\mu-\epsilon\dot{v}_\mu+\ \ldots\}. \quad . \quad . \quad (42)$$

Again, from Dirac (60) we obtain to the required order

$$\mathrm{F}_{\mu\nu}=e[1-(\boldsymbol{\gamma},\ \dot{\mathbf{v}})]^{-\frac{1}{2}}\{-\epsilon^{-3}(v_\mu\gamma_\nu-v_\nu\gamma_\mu)-\tfrac{1}{2}\epsilon^{-1}(\dot{v}_\mu v_\nu-\dot{v}_\nu v_\mu)+\ \ldots\}. \quad . \quad (43)$$

Dirac has also shown that the three-dimensional area $d\mathrm{S}$ is given by

$$d\mathrm{S}=d\Sigma[1-(\boldsymbol{\gamma},\ \dot{\mathbf{v}})]ds_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

where $d\Sigma$ is an element of the spherical normal cross-section of the electron tube.

The contribution of the electron tube to I_e is

$$\frac{1}{8\pi}\iiint\lambda^\mu(a^\nu+\mathrm{A}^\nu)(f_{\mu\nu}+\mathrm{F}_{\mu\nu})d\mathrm{S},$$

$$=\frac{1}{8\pi}\iiint\lambda^\mu\mathrm{A}^\nu\mathrm{F}_{\mu\nu}\,d\mathrm{S}+\frac{1}{8\pi}\iiint\lambda^\mu a^\nu\mathrm{F}_{\mu\nu}d\mathrm{S}+\ \ldots, \quad . \quad . \quad . \quad (45)$$

where the terms not written down vanish with ϵ. A^ν and $\mathrm{F}_{\mu\nu}$ refer to the electron's own field, a^ν and $f_{\mu\nu}$ to the field of the remaining electrons.

With the aid of (42), (43), and (44) we see that the first term is

$$-\frac{e^2}{8\pi}\iiint\epsilon^{-3}\lambda^\mu\{v^\nu-\epsilon\dot{v}^\nu+\}\{(v_\mu\lambda_\nu-v_\nu\lambda_\mu)+\tfrac{1}{2}\epsilon(\dot{v}_\mu v_\nu-\dot{v}_\nu v_\mu)+\ \ldots\}d\Sigma\,ds_0$$

$$=-\frac{e^2}{8\pi}\iiint\epsilon^{-3}\lambda^\mu\{v^\nu(v_\mu\lambda_\nu-v_\nu\lambda_\mu)-\epsilon[\dot{v}^\nu(v_\mu\lambda_\nu-v_\nu\lambda_\mu)$$

$$-\tfrac{1}{2}v^\nu(\dot{v}_\mu v_\nu-\dot{v}_\nu v_\mu)]+\ \ldots\}d\Sigma\,ds_0$$

$$=\ \frac{e^2}{8\pi}\iiint\{\epsilon^{-3}\lambda^\mu v^\nu v_\nu\lambda_\mu-\tfrac{1}{2}\epsilon^{-2}\lambda^\mu\dot{v}_\mu+\ \ldots\}d\Sigma\,ds_0,$$

the other terms vanishing in consequence of the relations

$$(\boldsymbol{\lambda},\ \mathbf{v})=0,\quad (\mathbf{v},\ \dot{\mathbf{v}})=0.$$

On performing the integration over the cross-section Σ the second term vanishes and the first term gives

$$-\frac{1}{2}\int\frac{e^2}{\epsilon}ds_0.$$

In the second term on the right-hand side of (45) we need only retain the first term in the expansion of $\mathrm{F}_{\mu\nu}$. A simple calculation shows that it gives rise to

$$\frac{1}{8\pi}\iiint\frac{e}{\epsilon^2}\lambda^\mu a^\nu v_\nu\lambda_\mu\,d\Sigma\,ds_0=-\frac{e}{2}\int a^\nu v_\nu\,ds_0.$$

The total contribution of a single electron tube is, therefore,

$$\frac{1}{2}\int\left\{-\frac{e^2}{\epsilon}-e(\mathbf{a},\,\mathbf{v})\right\}ds_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

which is the result (21).

We shall now dispose of I_s, namely, the contribution of the infinite sphere between the instants t_1 and t_2 where

$$t_1=-T+R,\ t_2=T+R, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

R ultimately becoming infinite.

In this case it is easily seen that $\lambda_0=0$, and $(\lambda_1,\,\lambda_2,\,\lambda_3)$ are the direction cosines of the normal to the sphere. A simple calculation shows that

$$I_s=\frac{1}{8\pi}\iint\int_{t_1}^{t_2}(\boldsymbol{\lambda},\,\phi\mathbf{E}+[\mathbf{a},\,\mathbf{H}])dS\,dt, \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

where [ ] denotes vector product, and we have returned to the usual notation.

Now

$$\phi\mathbf{E}+[\mathbf{a},\,\mathbf{H}]=-\phi\left(\text{grad}\,\phi+\frac{\partial\mathbf{a}}{\partial t}\right)+[\mathbf{a},\,\text{rot}\,\mathbf{a}]$$

$$=-\frac{\partial}{\partial t}(\phi\mathbf{a})-\frac{1}{2}\text{grad}\,\phi^2-\mathbf{a}\,\text{div}\,\mathbf{a}+[\mathbf{a},\,\text{rot}\,\mathbf{a}]. \quad . \quad (49)$$

If we write

$$\mathbf{P}=-\mathbf{a}\,\text{div}\,\mathbf{a}+[\mathbf{a},\,\text{rot}\,\mathbf{a}], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

and

$$P_{xx}=-\tfrac{1}{2}(a_x{}^2-a_y{}^2-a_z{}^2),\ P_{xy}=-a_xa_y,\ P_{xz}=-a_xa_z, \quad . \quad . \quad . \quad (51)$$

it is easily verified that the x-component of the vector $\mathbf{P}$ is

$$P_x=\frac{\partial P_{xx}}{\partial x}+\frac{\partial P_{xy}}{\partial y}+\frac{\partial P_{xz}}{\partial z} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

We have thus shown that the integrand on the right-hand side of (48) can be expressed as the sum of differential coefficients with respect to a space or time coordinate of homogeneous quadratic functions of the electromagnetic potentials. The differential coefficients with respect to the time coordinate may be integrated and expressed in terms of the values of the potentials at the limits, which, of course, refer to the bounding light cones.

Now P_{xx} will contain terms of the type

$$\frac{f(t_1,\,t_2)}{r_1r_2},$$

where

$$t_1=t-r_1,\ t_2=t-r_2,\ r_1{}^2=(X-x_1)^2+(Y-y_1)^2+(Z-z_1)^2,\ \text{etc.}, \quad (53)$$

and (x_1, y_1, z_1) is the position of electron 1 when it emits the disturbance which reaches the point (X, Y, Z) of the boundary at the instant t. (x_1, y_1, z_1) etc., are assumed to be small in comparison with (X, Y, Z).

We are interested only in terms which are of order R^{-2}, for the contributions of the remainder vanish when $R\to\infty$. To this order, we may, with the help of (53) write the differential coefficient with respect to x of the term $\frac{f}{r_1r_2}$ as

$$\left(\frac{\partial f}{\partial t_1}\frac{\partial t_1}{\partial X}+\frac{\partial f}{\partial t_2}\frac{\partial t_2}{\partial X}\right)\Big/R^2=-\left(\frac{\partial f}{\partial t_1}\frac{\partial t_1}{\partial t}+\frac{\partial f}{\partial t_2}\frac{\partial t_2}{\partial t}\right)\frac{\partial R}{\partial X}\Big/R^2=-\frac{\partial f}{\partial t}\frac{\partial R}{\partial X}\Big/R^2. \quad (54)$$

This term can now be integrated with respect to t and disposed of in the same way as the term which was an exact differential coefficient with respect to t. We have thus shown that I_s ultimately depends only on conditions on certain sections of the two bounding light cones, and, therefore, need not be taken into consideration when a finite portion of the world line of an electron is displaced.

References.

Born, M., *Ann. d. Phys.* xxviii. pp. 571–584 (1909).
Frenkel, J., *Z. Phys.* xxxii. 2, pp. 518–534 (1925).
Fokker, A. D., *Z. Phys.* lviii. pp. 386–393 (1929).
Wisskopf, V., *Phys. Z.* xxxiv. pp. 1–24 (1933).
Lewis, T., Phil. Mag. xx. pp. 1000–1025 (1935).
Milne, E. A., Proc. Roy. Soc. A, clxv. pp. 313–354 (1938).
Dirac, P. A. M., Proc. Roy. Soc. A, clxvii. pp. 148–169.
Pryce, M. H. L., Proc. Roy. Soc. A, clxviii. pp. 389–401 (1938).
Scherrer, W., *Helvetica Phys. Act.* xi. pp. 249–269 (1939).

L. *Haidinger's Rings in Non-Uniform Plates.*

By Sir C. V. Raman and V. S. Rajagopalan *.

[Received July 22, 1939.]

[Plates XIV. & XV.]

1. *Introduction.*

As is well known, the interference figures exhibited by transparent plates are of two distinct kinds, the typical examples of the same being the well-known rings of Newton and of Haidinger respectively. The figures of the Newtonian variety are exhibited by a plate of non-uniform thickness, and being dependent on the variations of thickness are localized at the surface of the plate itself. Haidinger's rings, on the other hand, are exhibited by a plane-parallel plate, and as they owe their origin to the varying inclinations of the rays of light, are in focus at infinity. The circumstances favourable for the observation of the two types of interference figures being thus very divergent, it has been generally assumed that the two phenomena are mutually exclusive, and in particular that the Haidinger rings can only be observed in plates which are exactly or very nearly plane-parallel. It will be shown in the present paper that this is not correct, and that with suitable arrangements we may observe the interferences of varying inclination in plates which are far from being either plane or parallel. Further, it is possible to follow the transition from interference figures of the Haidinger class to those of the Newtonian variety in plates of variable thickness. These facts are important for a correct appreciation of the theory of interferences in transparent plates.

2. *Some Theoretical Considerations.*

In instruments such as the Fabry-Perot etalon or the Lummer-Gehrcke plate giving very sharply defined interference fringes, perfect plane-parallelism of the optical surfaces is essential. When it is a question, however, merely of observing the interferences of varying inclination in a plate, such optical perfection is unnecessary ; the lack of it may be set off by a reduction in the aperture of observation. We shall now calculate the aperture required for observing the Haidinger rings in a uniform plate of thickness t and refractive index μ. If light of wave-length λ is incident at an angle i on the plate and is refracted into it at an angle r, the angular separation of the successive rings is $\mu\lambda \cos r/t \sin 2i$. In order that the successive rings should be resolved as seen through a circular aperture of diameter a in a direction inclined at an angle i to the normal, the angular separation of the rings should evidently exceed $\lambda/a \cos i$. This condition is always satisfied provided $a > 2t \tan \alpha$ where α is the critical angle for the material of the plate. If, however, we wish to observe the rings only in the vicinity of the normal to the plate, it is sufficient if $a > 2t \tan r$ where r corresponds to the outermost ring observed. An aperture of about twice the thickness of the plate is thus necessary for observing the complete Haidinger pattern, but a much smaller aperture than this would be sufficient to enable the first few rings to be seen with full definition. Using such a restricted aperture, non-uniformity in the plate is of no importance, provided the effective retardation produced by the plate does not alter by more than a fraction of the wave-length of light within the area actually employed in observing the rings.

---

* Communicated by Sir C. V. Raman, M.A., F.R.S.

A restriction of the effective area of the plate by a small circular aperture should thus enable the Haidinger rings to be observed even when the thickness is variable. It should, however, be remarked that when such variation is large the aperture used would have to be correspondingly small. Besides reducing the illumination available, this would have the result of reducing or destroying the visibility of the outer and more closely spaced rings in the pattern. In such a case it is advantageous to use, instead of a circular aperture, a *narrow rectilinear slit*, the same being placed parallel to the contour lines of constant thickness in the plate. In other words, we utilize a narrow strip of the surface which runs perpendicular to the direction in which the thickness of the plate changes most rapidly at the given point. When such a strip is narrow and not too long, the area of the plate which it covers could be regarded as a region of uniform thickness. The visibility of the Haidinger rings in the direction perpendicular to the slit would in these circumstances be reduced by diffraction to a negligible value. On the other hand, however, the visibility of the rings in the direction parallel to the strip would retain the full or maximum value. Thus, by using an orientated and sufficiently narrow slit, we should be able to observe the Haidinger pattern for a plate of variable thickness satisfactorily though, of course, only a strip of the pattern parallel to its diameter and running through its centre would be visible.

3. *Some Experiments and Observations.*

From the foregoing discussion, it is clear that the aperture of observation has to be restricted at or very near the surface of the plate itself if the interference figures of varying inclination are to be observable with a non-uniform plate. If we are viewing the Haidinger rings by transmission, this is a simple matter. The visibility of the transmitted system of rings is, however, low and they are therefore unsuitable for our present purpose. It is necessary to use the reflected system of rings, but the well-known methods of illuminating the plate with a reflector held at 45° or alternatively with a concave mirror containing an observation hole are unsuitable, as they do not enable the aperture to be restricted sufficiently close to the surface of the plate. The following arrangements are found more satisfactory. The image of a mercury lamp followed by a green ray filter is focused by a condensing lens on the white surface of a thin Bristol-board, the rear surface of the same being blackened. The transparent plate under observation is held in front of the illuminated area, and the light diffused forwards by the board and then reflected backwards by the surfaces of the plate is viewed through a small circular aperture made in the former. The distance between the plate and the illuminated screen containing the aperture may then be varied from zero, that is from actual contact between them, up to any desired large separation. Sufficient intensity is obtained with these arrangements to observe and photograph the interferences even when the aperture is only a fractional part of a millimetre in diameter. With the screen and plate in actual contact no light enters the aperture, but a separation of a millimetre or two is sufficient to enable the reflected light to pass through. The area of the plate utilized is then not much larger than that of the aperture itself.

If, instead of a small circular aperture, it is desired to employ a rectilinear slit, the same arrangements are suitable, except that some provision is desirable for adjusting the width of the slit and altering its orientation relatively to the plate under observation. This is easily arranged by cutting a slit of suitable shape in the illuminating screen and covering the same by two safety razor blades, leaving a fine slit between them. The screen is fixed on a rotating mount so that it can be turned round carrying the slit with it at the centre.

The efficiency of the arrangements described above is demonstrated by the facility with which they enable the Haidinger rings to be seen under the widest variety of conditions. Neither planeness nor parallelism in a transparent plate is required for this purpose. Fig. 1 (Pl. XIV.) represents the Haidinger rings photographed with a fragment of a thin spherical bulb blown from soft glass, the aperture of $\frac{1}{4}$ mm. diameter in the illuminated screen being held very close to its surface. Fig. 2 (Pl. XIV.) shows the same fragment held 5 centimetres from the hole and indicates that the glass, besides being spherical in shape, is far from being uniform in thickness. The arrangements described, in fact, enable the Haidinger rings to be observed and photographed with any piece of glass to be found in the laboratory such as a microscope slide or a cleaned-up photographic negative or even an ordinary window-pane.

The degree of perfection with which the Haidinger rings are exhibited by a non-uniform plate depends on the aperture used and its distance from the plate, as also upon the thickness of the latter and its variability. With the aperture sufficiently close to the plate, the relation between the width of the aperture and the gradient of thickness in the plate is evidently the determining factor. The smaller the aperture is in relation to the thickness-gradient, the more perfectly would the interference figures of equal inclination be formed. It must be remarked, however, that the observing aperture cannot be indefinitely reduced in area. For, as has already been explained, diffraction effects would come into play and reduce the visibility of the rings, thereby setting a lower limit to the size of the aperture that can be used. On the other hand, there is also an upper limit to the size of aperture which can be employed; this corresponds to the case in which the aperture covers a variation of several wave-lengths in the effective thickness of the plate. The Haidinger rings corresponding to the different thicknesses would then be superposed, thereby reducing their visibility to zero. There is thus an optimum width of aperture for any particular case which makes the visibility of the Haidinger rings a maximum. It should be noted that this visibility is not necessarily the same in different parts of the field of observation. In general, when a very small aperture is used in order to reduce the effect of the thickness-gradient, the visibility of the rings would be a maximum at the centre of the field and would diminish from the centre outwards. It may also, in general, be unsymmetrical about the centre of the field in the direction of the gradient of thickness for reasons which will presently become apparent.

When the thickness of the plate is so great and so variable that the use of a circular aperture becomes ineffective, the method of the orientated slit already described becomes specially useful. Fig. 3 (Pl. XIV.) shows the Haidinger pattern photographed by the latter method with a glass plate approximately 3·5 millimetres thick, but so non-uniform that distant objects exhibit by reflexion in it two distinct images of varying separation. It will be seen in the figure that a very large number of interferences are visible running at right angles to the length of the slit, while there are none to be observed in other directions. A distinct asymmetry in the visibility of the fringes about the centre of the field is also noticeable. With the method of the orientated slit, it is generally found that as the slit is turned round, the Haidinger interferences appear and disappear twice in every complete rotation, as is to be expected from the theoretical considerations already set out. In special cases, however, the visibility of the fringes varies in a more complicated manner, and these presumably correspond to cases in which the contours of equal thickness on the plate do not run straight but are strongly curved.

4. *Transition from Haidinger to Newtonian Interferences.*

It is clear that when the observing aperture is not near the plate, the form of the interferences is necessarily modified by the superposition of the effects of variable thickness and of varying inclination. In other words, it becomes possible to observe the transition from the Haidinger rings to the Newtonian variety of interference figure, and this is automatically secured as the plate under observation is moved away from the illuminated screen. The variations of path due to varying inclination of the rays then diminish, while, on the other hand, the variations of path due to varying thickness remain constant. Thus with the plate sufficiently distant from the aperture, we observe merely the Newtonian fringes, whereas with the plate sufficiently close to the aperture, we see the Haidinger rings. Intermediately we have a transition from one type to the other. The rapidity with which such transition occurs depends on the magnitude of variation of the thickness of the plate. The greater the variation of thickness in any particular part of the plate, the more closely should this area of the plate approach the aperture of observation before the rings of variable thickness change over to the Haidinger rings. On the other hand, in the areas of comparative uniformity of thickness, the transition occurs even when the plate is at considerable distances from the observing aperture.

Figs. 4, 5, and 6 (Pl. XV.) are the first three of a set of photographs taken to illustrate the foregoing remarks and were obtained with a microscope cover-slip 0·2 millimetre thick and 2 centimetres square, the distances from plate to aperture for the three cases being respectively 0·5 cm., 3 cm., and 6 cm. Separations larger than 6 cm. gave photographs very similar to fig. 6 on a smaller scale, while separations smaller than 0·5 cm. made no change from the pattern shown in fig. 4, apart from reducing the intensity of the first few rings. It will be noticed that fig. 5 shows a notable distortion and asymmetry of the rings and is clearly intermediate in type between fig. 4 and fig. 6. The small black dot appearing towards the centre of figs. 5 and 6 is the image of the observing aperture (0·5 mm. diameter) seen by reflexion at the surface of the plate. In fig. 6 it is seen to be smaller than the width of the interference fringes on which it is superposed, while in fig. 5 it appears larger in size and is on the edge of the first ring instead of at its centre. In fig. 4 it seems much larger and is very nearly central. The size of the aperture seen projected against the interferences of variable thickness as in fig. 6 gives us a direct measure

FIG. 3.

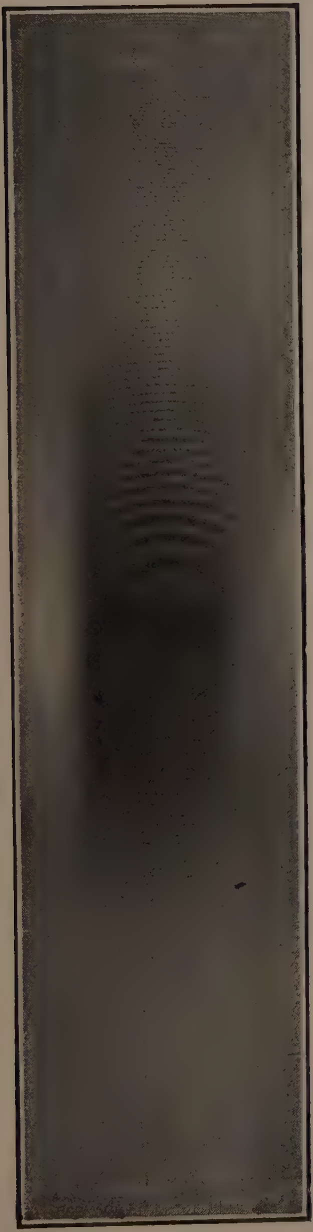


FIG. 1.

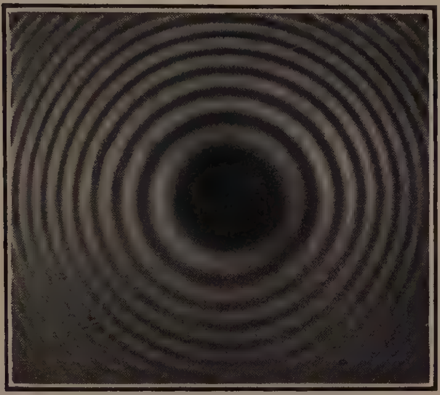
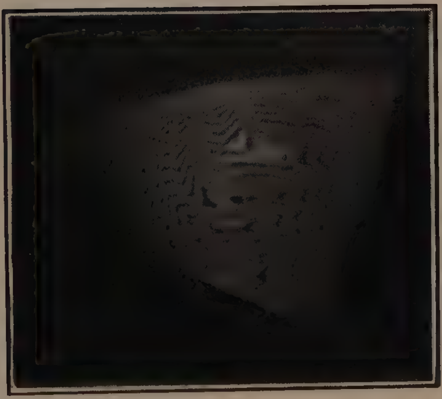


FIG. 2.



Interference patterns in non-uniform plates.

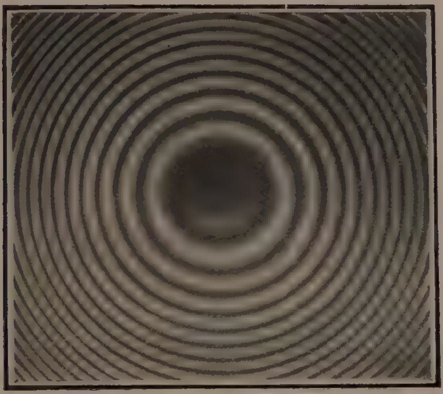


Fig. 4.

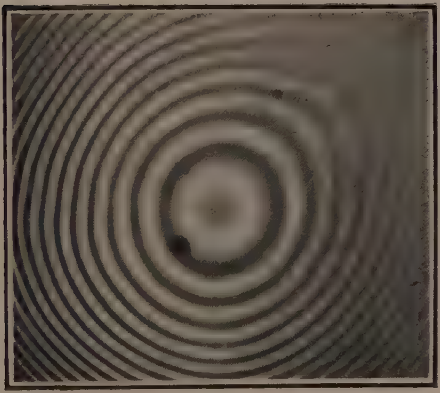


Fig. 5.

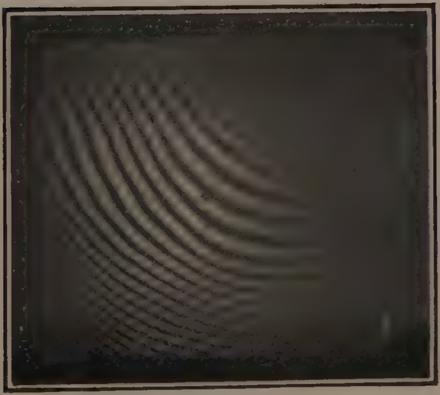


Fig. 6.

Interference patterns in non-uniform plates.

of its suitability for observing the Haidinger rings when held close to the plate as in fig. 4.

The configurations which the interference figures assume as the plate is moved away from the observing aperture necessarily depend on the circumstances of the particular case, namely, the shape of the plate and the form of the contours of constant thickness. We may, however, note some features of a general validity. In the Haidinger rings, the order of the interference *diminishes* from the centre outwards. On the other hand, in the Newtonian fringes, the order of the interference *increases* with increasing thickness. Thus, the superposition of the variations of path arising in the two different ways would result in shifting the position of the interferences variously in various directions. Transverse to the gradient of thickness, the position of the interferences of varying inclination would be unaffected, while along the gradient the curves would be crowded up in one direction and go wider apart in the other. These effects are clearly seen in fig. 5. It will be noticed in this figure that the visibility of the fringes is the maximum in the direction in which they are closer together, while it becomes relatively small in the direction which the fringes are further separated, thus showing a marked asymmetry.

5. *Variation of Focus and Visibility.*

Since the Haidinger rings are formed at infinity while the fringes of variable thickness are localized on the plate itself, a transition from one to the other involves a change of the plane at which the fringes are formed most clearly. With a very small aperture covering the lens of the camera, the effect of this variation of focus is not great, particularly as the displacement of the plate away from the aperture in the transition from the Haidinger to the Newtonian interferences automatically secures part of the adjustment of focus necessary. Indeed, figs. 4, 5, and 6 (Pl. XV.) were all photographed without altering the setting of the camera for infinity. Nevertheless, the question of focus is of some importance, as it is closely connected with the theory of the formation of the Haidinger rings with a non-uniform plate, and especially with the variations of their visibility in different parts of the field.

The simplest case to consider theoretically is that of a wedge-shaped plate. A parallel beam of light falling on it would be reflected from back and front in two different directions, and these reflected beams would form images at different points in the focal plane of a camera set for infinity. They would therefore be incapable of interfering with each other. But if the aperture be limited by a slit parallel to the contours of constant thickness, the images would spread out by diffraction, and if they overlap sufficiently interference would become possible. The necessity for restricting the aperture in order to obtain visible interferences is thus readily understood. It is also clear that too great a restriction of aperture would again result in destroying the visibility of the interferences in the plane containing the angle of the wedge. For light falls on the plate over a whole range of incidences, and the diffraction by a very narrow slit would result in superposing the interference patterns due to light incident at widely different angles and thus producing uniformity of illumination instead of any observable fringes.

If we consider a thin pencil of rays incident on a wedge-shaped plate and reflected from its back and front, it will be noticed that the two reflected beams *diverge* from each other if the incidence is on the side of the normal towards the thicker side of the wedge, and that they *converge* towards each other if the incidence is on the thinner side. The region where the two beams meet and interfere would therefore be in front of the plate in one case and behind the plate in the other. This is clearly the origin of the asymmetry in the visibility of the Haidinger rings on either side of the centre which has already been remarked upon.

Summary.

By the use of an opaque screen which diffuses light and which can be brought very close to the transparent plate and observing the latter through an aperture of suitable size and shape in the screen, it is possible to see and photograph the Haidinger rings by reflexion in plates which are neither plane nor parallel. The most efficient form of aperture is a narrow slit which is orientated parallel to the contours of equal thickness on the plate ; the use of this enables a diametral section of the Haidinger pattern containing a very large number of perfectly defined interferences to be observed even with relatively thick and non-uniform plates. The use of a small circular aperture instead of a slit enables the complete rings to be observed, but in this case the limitation of aperture impairs the definition of the outer and more closely spaced rings. The transition from the Haidinger to the Newtonian variety of interferences may also be followed using the diffusing screen with a small circular aperture and gradually removing the plate under observation to a sufficient distance from it.

References.

Rayleigh, Phil. Mag. xii. p. 489 (1906).
Chinmayanandam, Proc. Roy. Soc. A, xcv. p. 176 (1918).
Walker, ' Analytical Theory of Light,' pp. 62–68.
Max Born, *Optik*, pp. 120–122.
O. Lummer ' Muller-Pouillets Lehrbuch der Physik,' 11th edition, 2nd vol. 1st part, pp. 698–706.

LI. *Note on " The Theory of Light-scattering."*

By R. S. KRISHNAN *.

[Received February 19, 1940.]

IN a recent number of this Journal (1) there appeared a paper by S. Parthasarathy entitled "On the Theory of Light-scattering." It is stated therein that the paper was based on some experimental work published by me in a series of papers in the 'Proceedings of the Indian Academy of Sciences.' I would like to make the following remarks with reference to Dr. Parthasarathy's paper :—

1. If a parallel beam of light is incident on a lens of aperture F/4·5, the semi-angle of convergence θ of the extreme rays at the focus is nearly equal to 1/9 of a radian. Hence according to the formula given by Gans (quoted in his paper) if ρ_v is equal to 0·4 per cent. the calculated value of ρ_h will be equal to 4·4 as given in Table II. on page 151. If, on the other hand, instead of using a parallel beam of light, one uses an illuminated slit as the source of light (as was used in my experiments (2)), the rays are brought to a focus by the lens at a distance of twice the focal length and the semi-angle of convergence θ for an aperture of F/4·5 will be equal to 1/18 of a radian. The calculated value of ρ_h for such a case comes to about 1·25 when ρ_v is equal to 0·4 per cent., *i. e.*, the correction to be applied to the observed value is smaller still.

2. As has been clearly stated in my papers (2), I had consistently used a long focus short aperture lens (F=30 cm. and aperture=4 cm.) so as to reduce the convergence correction to a minimum, and an illuminated slit as the source of light in order to get a well-defined track inside the scattering medium. The angle of convergence was of the order of 4°, which was further reduced to 3° inside the medium by the liquid in the rectangular cell. It can be seen from theory that for such a small angle of convergence the correction for ρ_v is equal to 0·015 per cent., which is extremely small. If the observed value of ρ_v is 0·4 per cent., the value of ρ_h calculated from Gans's theory comes to about 1·08 per cent., which is not appreciably different from unity. The observed values of ρ_h are far removed from unity and therefore cannot be accounted for on the basis of the finite angle of convergence of the incident beam. It should

---

* Communicated by the Author.

be added that the optical arrangement was kept unaltered during experiments with different liquid mixtures in contradiction to what Parthasarathy suggests in his paper. Parthasarathy postulates not only high values for the angle of convergence, but also different values for different mixtures. These assumptions do not bear any relation to the actual conditions of the experiment.

3. In one of my papers (3) I had reported that I could not detect any difference in the values of ρ_u, ρ_v and ρ_h when the aperture of the lens was *reduced* from 4 to 1 cm. (*i. e.*, the angle of convergence was reduced from 3° to 0·75°), showing thereby that for such a small angle the convergence correction was so small that its neglect was justified. On page 153 of his paper, Dr. Parthasarathy misquotes the above sentence to suit his arguments. He says " That I did not find any difference when the aperture was *increased* four times from the original value."

4. The formula for the convergence correction given by Gans and Ananthakrishnan is applicable only to the case of scattering of light by molecules and very small particles in suspension. In the case of large particles and molecular aggregates, as in protein solutions and liquid mixtures very near the critical solution temperature, ρ_h is definitely greater than unity because of the appreciable size of the scattering particles in relation to the wave-length of light. From *a priori* considerations one can say that the effect of convergence in such cases will be to decrease the value of ρ_h and not to increase it.

5. It is well known that in protein solutions aggregation of micelles takes place at the isoelectric point. The recent optical investigations of Subba Ramaiah (4) show that the $\rho_h - \rho_v$ curve for such solutions at the isoelectric point is similar to the $\rho_h - \rho_v$ curve for liquid mixtures at the critical solution temperature.

References.

(1) S. Parthasarathy, Phil. Mag. xxix. p. 148 (1940).
(2) R. S. Krishnan, Proc. Ind. Acad. Sci. (A) i. p. 915 ; ii. p. 221 (1935).
(3) R. S. Krishnan, Proc. Ind. Acad. Sci. (A) i. p. 919 (1935).
(4) K. Subba Ramaiah, Proc. Ind. Acad. Sci. (A) v. p. 128 (1937).

Trinity College, Cambridge.
22nd February, 1940.

LII. *Notices respecting New Books.*

Electro-acoustics. By Prof. Erwin Meyer. [Pp. xi+117.] (London: G. Bell & Sons, Ltd. Price 10*s.* net.)

This is a useful and welcome addition to the existing literature on acoustics. It deals mainly with recent developments in this field and overlaps very little with other books on acoustics which have been published during the last few years.

On the theoretical side, electro-acoustics demands a considerable knowledge of both acoustics and electrical theory, and it is probably safe to say that only a comparatively few specialists will be equally conversant with both aspects. One has only to call to mind, however, a few of the many ways in which electro-acoustic developments affect present day life—*e. g.*, loudspeakers, microphones, sound recording and reproduction—to realize that on the practical side the subject is of interest to a much wider circle. A practical interest often leads to a desire to know something about the theoretical side of a subject; the present book will be found useful and stimulating to anyone wishing to take a first step in this direction in electro-acoustics.

Although the subject-matter of the book originally constituted a series of five lectures to electrical engineers, the book will be useful to other classes of reader. Teachers and students who are occupied more with the academic than with the applied side of acoustics, will find the book of value as giving them a better acquaintance with developments on the technical side and with some of the problems still not fully solved.

The style of the book is very concise and the author passes rapidly from one point to the next; in consequence a good deal of ground is covered. The reader who is interested in the subject will no doubt often regret that lack of space precluded a fuller exposition of many of the topics briefly mentioned, but since this would necessitate a much larger, and therefore more expensive, book, he will no doubt be grateful for a compact book covering so much ground.

G. W. B.

Physico-chemical Methods. By Prof. J. Reilly and Prof. W. N. Rae. [Two vols. pp. xv+686 and ix+580.] (London: Methuen and Co., Ltd., 1940. Price £4 4*s.* net.)

The authors of this comprehensive and admirable treatise on physico-chemical methods have, in their own words, resisted "the temptation to convert the work into an unwieldy encyclopædia." Nevertheless, the work has something of an encyclopædic character. It is impossible to turn its pages without lighting on some fact, hint, or working tip which at once arouses interest. Of its value there can be no question—the fact that it has reached a third edition speaks for itself.

This edition of nearly 1300 pages (compared with the 800-odd pages of its predecessor) is now conveniently divided into two volumes, one dealing with measurement and manipulation, one with physical properties. Though, in these days of increased expense and of scant leisure, we may regret the increase of price and of size, we have to confess that the advantages of the new edition far outweigh the disadvantages, for the authors have added much of value, and have neglected little of importance.

The range of topics is immense, and much of the information is necessarily given in condensed, though not necessarily in sketchy, form. The authors have, most wisely, given a carefully-selected bibliography of books and papers under the title " suggestions for further reading " which is classified under the different chapter-headings.

It is the first duty of a reviewer to give his readers some notion of the contents of the book under review, but it would be an impossible task to list all the topics which the authors have handled. It must suffice to say that the first volume is concerned with basic measurements, with the methods of reduction of observations, with laboratory arts and manipulations, with high pressure technique and thermometry, and with the measurement of physical properties such as surface tension, viscosity, elasticity, solubility, molecular weight, partition coefficients and so forth.

The second volume is concerned with the handling of gases and vapours, with separation processes, with optical and electro-chemical measurements, with thermionic tubes, dielectrics, radio-activity, hydrogen-ion concentration and colloids.

Truly a comprehensive list, and yet one which is by no means exhaustive of the contents of the book.

Where so much has been given, it would seem ungracious to ask for more, and, indeed, the reviewer would not wish to see the book expanded beyond its present limits. But, it is possible, by cutting out some of the more elementary demonstrations, to find room for the elucidation of problems which are often a source of very real trouble. For example, in the section dealing with graphical methods, is it not possible to omit, or compress, the discussion of rectangular coordinates and of the simpler curves, and thus to find room for the discussion of graphical methods for the determination of the constants in such important equations as

$$y=a+bx^n \quad \text{or} \quad y=a(1-bx)^n,$$

where y and x are experimentally observed quantities ?

And how may one quickly, and fairly accurately, draw a tangent to a curve ?

The book is admirably produced ; it is a necessary addition to the library of any well-equipped laboratory ; and the authors and their collaborators are to be heartily congratulated on the production of a work which is certainly assured of success.

A. F.

The Philosophy of Physical Science. By Sir A. S. EDDINGTON. [Pp. i–ix+1–230.] (Cambridge University Press, 1939. Price 8*s.* 6*d.* net.)

THIS book contains so many points of interest and challenges to current scientific thought that an adequate review of it would be as long as the original work itself : a short notice can only touch on a few of the questions raised. A starting-point may be found in the notion, now familiar to quantum physicists, that the result of an experiment is conditioned by the methods of conducting the experiment. Eddington's view is a generalization of this idea. He would say that a great part of physical science reflects not so much the properties of an objective world, but the forms of thought of the physicist himself. The inescapable and universal character of certain elements of physical knowledge is due to their subjective nature. Hence an examination of this knowledge itself should detect these elements, an enquiry which Eddington terms the "epistemological method" in science. It is asserted that the laws and constants of nature, such as the law of gravitation or the celebrated number 137 in quantum theory, are precisely these elements and are therefore discoverable by *a priori* analysis of our methods of acquiring knowledge of the physical world. Of these conditioning "forms of thought," the principal are the concepts of analysis, of structure, and of existence. The first perpetually urges us on to the construction of a universe consisting of the relations of units intrinsically alike, the second to the group-theory of these relations, whilst the third is an illusion which should have no part in physics. As the chief example of the power of the epistemological method, Eddington gives a general account of his calculation of the (finite) cosmical number which is the total number of elementary particles in the universe. The book concludes with a discussion of the reasons for admitting that there are other minds besides the investigator's own, and thus to an escape from the solipsism to which this type of philosophy seems to be condemned.

Many readers will probably lay down this book with a profound sense of disappointment. They will have looked for an exhaustive and reasoned argument leading up to the conclusion that the laws and constants of nature are subjective. Instead they are confronted by the *fait accompli* of this result, supported by an analogy of the fisherman and his net. Moreover, the case for subjectivism is weakened by the admission (on p. 63) that there are "uniformities intrinsic in the world around us." Such uniformities would present the same universal and inescapable character as does, for instance, the law of gravitation. No entirely satisfactory criterion for distinguishing epistemological laws from these intrinsic uniformities is to be found in Eddington's work. Again, the detailed calculation of the cosmical number given in "Relativity Theory of Protons and Electrons," is well known to depend on the use of the Einstein-de Sitter class of universes. But universes of this class do not, in fact, describe the observed recession and distribution of the spiral nebulæ. If the hyperbolic universes, which do fit the observations, are chosen instead, the cosmical number is not finite. Thus Eddington's cosmical number is obtained less by an epistemological analysis of our knowledge of the physical universe, than by the introduction of the "physical" hypothesis, arbitrary from the standpoint of logic and epistemology, that quantum theory and general relativity are to be united through the Einstein-de Sitter universes. G. C. McV.

The Principles of Electrochemistry. By Duncan A. MacInnes. [Pp 478.] (Reinhold Publishing Corporation, New York, 1939. Price $6.)

This book has been written by a well-known worker in the field of electrochemistry, and the author says in his preface that he has written it to satisfy an inner urge. It will appeal to the student in that the subject is developed clearly from first principles and is presented not as a series of selected topics, but as a connected whole; and it will also appeal to the research worker and will help him to see how his own particular problems are related to the general electrochemical background.

Although the book deals with the principles of electrochemistry, it is mainly concerned with the interpretation of experimental data, and contains descriptions of experimental methods and the theory underlying them. Many tables of data have been included, and special care has been taken to give the best available values. Emphasis is thus laid on the present position of the subject rather than on its historical development. The treatment is based on thermodynamics, on which there is a clear introductory chapter, and it is of interest to find in a book published in America that the symbols used are very similar to those recommended by the Joint Committee of the Chemical Society, the Faraday Society, and the Physical Society (except that the Gibbs free energy is denoted by Z). The author points out that in a work of reasonable size the choice of subjects must be somewhat arbitrary. Most of the book is devoted to the properties of dilute aqueous solutions, but a chapter on the thermodynamics of non-aqueous solutions has been included. As well as excellent chapters on the thermodynamics of galvanic cells and associated topics such as liquid junctions, the determination and meaning of pH, oxidation-reduction potentials and potentiometric titrations, there is a discussion of the Debye–Hückel theory, transport phenomena, conductivity in aqueous, non-aqueous, and mixed solvents, and the ionization constants of organic acids and bases. The final chapters are devoted to a brief consideration of dielectric constants, electrokinetic phenomena, and irreversible electrode phenomena. Only well-established material is included, speculative matter being irrelevant in a work of this size. References to the essential original literature are given throughout.

The printing and diagrams are good, and there appears to be very few misprints. The book may be strongly recommended to all who are interested in electrochemistry and the properties of solutions. J. W. B.

Theoretical Electrochemistry. By N. A. McKenna. [Pp. 469.] (London: Macmillan and Co., Ltd., 1939. Price 15*s*.)

Although the title of this book is 'Theoretical Electrochemistry,' experimental methods have not been neglected, and there is an early chapter on fundamental electrical measurements. The first part, which deals adequately with the conductivity of solutions, plunges immediately into the Debye-Hückel-Onsager theory; then follows a discussion of transport phenomena, and then a chapter on the viscosities, diffusions, compressibilities, and surface tensions of electrolytes from the standpoint of interionic attraction. The second section of the book is introduced by a hardly adequate chapter on chemical thermodynamics; it is a pity that the symbols recommended by the Joint Committee have not been adopted. The Debye-Hückel theory and its applications are then treated in some detail, and then the theory of galvanic cells is given, including a section on the statistical theory of electrode processes. Finally, there is a brief account of overvoltage, of solvation, and of ionic equilibria.

The book is a valuable review of the subjects treated and includes further references to the literature. It is hardly suitable, however, for a student's text-book, as it assumes some previous knowledge and does not lead him gently enough up the more difficult slopes. It should prove useful as a reference book to those already partly familiar with the subject. The printing and diagrams are excellent. J. W. B.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

LIII. *Theory of Resonance in Microwave Transmission Lines with discontinuous Dielectric.*

By H. R. L. Lamont, M.A., B.Sc., Carnegie and Breadalbane Scholar in the University of Glasgow *.

[Received June 7, 1939.]

In a recent paper King⁽¹⁾ has described a method of measuring dielectric constants at ultra-high frequencies. The theory was developed of a stationary wave system on a pair of Lecher wires, these being surrounded for part of their length by a solid or liquid dielectric and for the remainder by air. The interesting mathematical result obtained seems to merit more exhaustive treatment, and to provide such is the object of the first part of this paper. Incidentally misconceptions by previous authors seem to remain uncorrected, and these will be pointed out. The second part deals with the effect of the introduction of a slab of dielectric in a hollow tube resonant line of the form investigated by Barrow⁽²⁾, Southworth⁽³⁾, and Carson, Mead and Schelkunoff⁽⁴⁾. The two characteristic types of wave form are studied, and it is shown how each is affected by the dielectric.

Part I.

Introduction.

A pair of parallel conductors by virtue of their mutual inductance and capacity act as a high-frequency oscillatory circuit, and may be tuned to resonance with an oscillator by means of a sliding bridge which varies the effective length of the wires. We assume here the ideal conditions of perfect reflexion at the bridge and loss-free conductors. Where the wires are in air the distance between successive resonance positions of the bridge then equals half the free space wave-length λ of the exciting wave.

If the wires are surrounded by a non-conducting material of dielectric constant ϵ the wave-length along the wires is reduced to a value λ_1, where $\lambda=\sqrt{\epsilon}\lambda_1$.

Theory.

(a) *Resonance Condition.*

The two systems represented in fig. 1 will be analysed.

(i.) represents a parallel wire system in air coupled at one end to a

---

* Communicated by Dr. John Thomson.

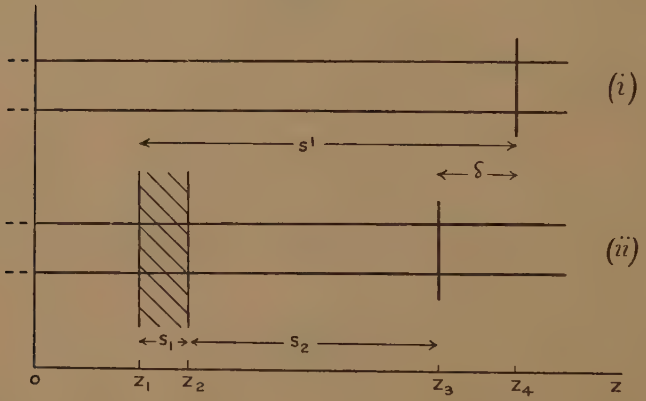
high-frequency oscillator and closed at the other by a movable bridge so placed that the system is in resonance.

(ii.) shows the same system with a section of the wires surrounded by a dielectric contained between planes perpendicular to the wires and supposedly infinite in extent. The frequency of the oscillator being constant, the bridge must obviously be moved to restore resonance.

Let an arbitrary origin be taken at O, and let z_1, z_2 be the abscissæ of the dielectric faces, and z_3, z_4 the abscissæ of the bridge positions in (ii.) and (i.).

Let zone 0 extend from $z=0$ to $z=z_1$,
zone 1 extend from $z=z_1$ to $z=z_2$,
zone 2 extend from $z=z_2$ to $z=z_3$.

Fig. 1.



The propagation constant in zones 0 and 2 is β, and in zone 1 is β_1, these being equal to $\frac{2\pi}{\lambda}$ and $\frac{2\pi}{\lambda_1}$ respectively. Solutions of the wave propagation equations for an angular frequency p give for the electric and magnetic intensities of a plane wave in Gaussian units

$$E=Ae^{i(pt-\beta z)}+Be^{i(pt+\beta z)},$$

$$H=\frac{c\beta}{p}\{Ae^{i(pt-\beta z)}-Be^{i(pt+\beta z)}\},$$

where c is the velocity of light. The first term is a wave progressing in the direction of z, and the second a wave in the direction of $-z$. A factor expressing the distribution in the (x, y) plane has been omitted as it plays no part in the succeeding work. The permeability factor μ is also omitted throughout, being assumed to be unity.

The same equations hold for the coaxial cable.

Denoting the above zones by subscripts, we get

$$\left.\begin{aligned}
E_0&=A_0e^{i(pt-\beta z)}+B_0e^{i(pt+\beta z)},\\
E_1&=A_1e^{i(pt-\beta_1 z)}+B_1e^{i(pt+\beta_1 z)},\\
E_2&=A_2e^{i(pt-\beta z)}+B_2e^{i(pt+\beta z)},\\
H_0&=\frac{c\beta}{p}\left(A_0e^{i(pt-\beta z)}-B_0e^{i(pt+\beta z)}\right),\\
H_1&=\frac{c\beta_1}{p}\left(A_1e^{i(pt-\beta_1 z)}-B_1e^{i(pt+\beta_1 z)}\right),\\
H_2&=\frac{c\beta}{p}\left(A_2e^{i(pt-\beta z)}-B_2e^{i(pt+\beta z)}\right).
\end{aligned}\right\}\quad\ldots\ldots\ldots\quad(1)$$

The boundary conditions at z_1 and z_2 require that E and H (which are, of course, in the (x, y) plane) should be continuous, *i. e.*,

$$\left.\begin{aligned}
A_0e^{-i\beta z_1}+B_0e^{i\beta z_1}&=A_1e^{-i\beta_1 z_1}+B_1e^{i\beta_1 z_1},\\
A_1e^{-i\beta_1 z_2}+B_1e^{i\beta_1 z_2}&=A_2e^{-i\beta z_2}+B_2e^{i\beta z_2},\\
A_0e^{-i\beta z_1}-B_0e^{i\beta z_1}&=n(A_1e^{-i\beta_1 z_1}-B_1e^{i\beta_1 z_1}),\\
n(A_1e^{-i\beta_1 z_2}-B_1e^{i\beta_1 z_2})&=A_2e^{-i\beta z_2}-B_2e^{i\beta z_2},
\end{aligned}\right\}\quad\ldots\ldots\quad(2)$$

where

$$n=\frac{c\beta_1}{p}\Big/\frac{c\beta}{p}=\frac{\beta_1}{\beta}=\sqrt{\epsilon},\quad\ldots\ldots\ldots\quad(3)$$

and is here the index of refraction of the medium. The substitutions $n=\sqrt{\epsilon}$ and $\beta_1=\sqrt{\epsilon}\beta$ are not made at present in order to allow comparison with the results in Part II., where these relations do not hold.

At the bridge end of the system, for perfect reflexion

$$A_2e^{-i\beta z_3}+B_2e^{i\beta z_3}=0.\quad\ldots\ldots\ldots\quad(4)$$

If we now choose O such that the incident and reflected waves in case (ii.) produce a node of electric intensity at O, then

$$A_0+B_0=0.\quad\ldots\ldots\ldots\quad(5)$$

Equations (2) and (5) now define the six constants in terms of any one of them. Solving in terms of A_0, we get

$$\left.\begin{aligned}
A_1&=\tfrac{1}{2}A_0e^{i\beta_1 z_1}\left\{\left(1+\frac{1}{n}\right)e^{-i\beta z_1}-\left(1-\frac{1}{n}\right)e^{i\beta z_1}\right\},\\
B_1&=\tfrac{1}{2}A_0e^{-i\beta_1 z_1}\left\{\left(1-\frac{1}{n}\right)e^{-i\beta z_1}-\left(1+\frac{1}{n}\right)e^{i\beta z_1}\right\},
\end{aligned}\right\}\quad\ldots\ldots\quad(6)$$

$$\left.\begin{aligned} A_2&=\tfrac{1}{4}A_0e^{i\beta z_3}\Bigg[(1+n)e^{-i\beta_1 z_2}\left\{\left(1+\frac{1}{n}\right)e^{i(\beta_1-\beta)z_1}-\left(1-\frac{1}{n}\right)e^{i(\beta_1+\beta)z_1}\right\} \\ &\qquad+(1-n)e^{i\beta_1 z_2}\left\{\left(1-\frac{1}{n}\right)e^{-i(\beta_1+\beta)z_1}-\left(1+\frac{1}{n}\right)e^{-i(\beta_1-\beta)z_1}\right\}\Bigg], \\ B_2&=\tfrac{1}{4}A_0e^{-i\beta z_3}\Bigg[(1-n)e^{-i\beta_1 z_2}\left\{\left(1+\frac{1}{n}\right)e^{i(\beta_1-\beta)z_1}-\left(1-\frac{1}{n}\right)e^{i(\beta_1+\beta)z_1}\right\} \\ &\qquad+(1+n)e^{i\beta_1 z_2}\left\{\left(1-\frac{1}{n}\right)e^{-i(\beta_1+\beta)z_1}-\left(1+\frac{1}{n}\right)e^{-i(\beta_1-\beta)z_1}\right\}\Bigg]. \end{aligned}\right\} \quad . \quad (7)$$

Substituting in (4), and reducing, we get the eliminant

$$n\cot\beta_1 s_1(\cot\beta z_1+\cot\beta s_2)+\cot\beta z_1\cot\beta s_2-n^2=0, \quad . \quad . \quad (8)$$

where z_2-z_1 and z_3-z_2 have been replaced by s_1 and s_2 respectively. The factor $\sin\beta_1 s_1 \sin\beta z_1 \sin\beta s_2$, which has been omitted from (8) for convenience, disposes of singularities. This equation expresses s_2 as a periodic function of z_1, and shows that the reduction in resonant length produced by the dielectric slab is a function of the absolute position of the slab. The result is somewhat unexpected, as a shift equal to the "optical path difference" between the dielectric and an equal thickness of air would seem plausible, and has indeed been assumed by some authors. It must be remembered, however, that in the optical problems where such a result is true the interference is between two separate travelling waves. Here the interference is between an incident wave and the fractions of it reflected at z_1, z_2, and z_3.

For case (i.), since O and z_4 are nodes, $\beta z_4=\nu\pi$, where ν is an integer, and if $z_1+s'=z_4=\dfrac{\nu\pi}{\beta}$ defines s' (8) can be rewritten

$$n\cot\beta_1 s_1(\cot\beta s'-\cot\beta s_2)+\cot\beta s'\cot\beta s_2+n^2=0. \quad . \quad . \quad (9)$$

This is the equation which King⁽¹⁾ has derived *. He has used the admittance expressions of the theory of the transmission line, which, although producing equation (9) more simply than the above, throw no light on the cause of the variation in resonant length with dielectric position, or the distribution of resultant amplitude and phase along the line. The various relations derived above enable these problems to be studied.

If δ is the shift of the bridge required to restore resonance after the insertion of the dielectric, s_2 can be replaced by $s'-(s_1+\delta)$. The graph of δ against s' as given by (9) with this modification is of the form shown in fig. 2.

---

* With the exception that King has introduced a term expressive of the effect of a thin retaining wall on the faces of the dielectric.

It can be shown by differentiation that the abscissæ of the turning points of this curve are given by

$$\tan \beta s' = \frac{1}{n} \tan \frac{1}{2} \beta_1 s_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

and

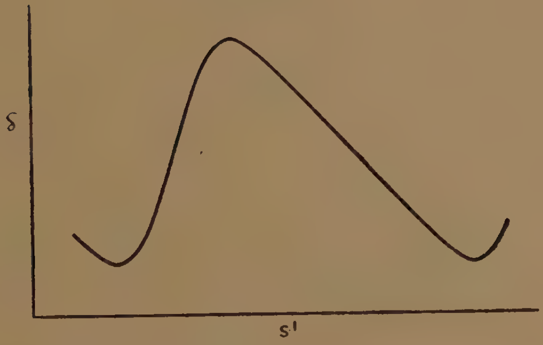
$$\tan \beta s' = -\frac{1}{n} \cot \frac{1}{2} \beta_1 s_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and the corresponding values of δ by

$$\tan \frac{1}{2} \beta (s_1 + \delta) = \frac{1}{n} \tan \frac{1}{2} \beta_1 s_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$\tan \frac{1}{2} \beta (s_1 + \delta) = n \tan \frac{1}{2} \beta_1 s_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Fig. 2.



Which of equations (12) and (13) represents $\delta_{\max}$ and which $\delta_{\min}$ depends on the sign of $\tan \frac{1}{2}\beta_1 s_1$, *i. e.*, upon the thickness s_1 of the slab. Since $n>1$, it is evident that when $\left[\frac{2s_1}{\lambda_1}\right]$* is zero or an even integer (12) gives $\delta_{\min}$ and (13) $\delta_{\max}$; for $\left[\frac{2s_1}{\lambda_1}\right]$ an odd integer (12) gives $\delta_{\max}$ and (13) $\delta_{\min}$ †.

From (10) and (12)

$$\begin{aligned} \tan \beta s' &= \tan \tfrac{1}{2}\beta(s_1+\delta) \\ &= \tan \tfrac{1}{2}\beta(s'-s_2), \\ \therefore \quad \beta s' &= \tfrac{1}{2}\beta(s'-s_2)+\nu\pi, \end{aligned}$$

i. e.,

$$s'+s_2=\nu\lambda,$$

---

* $[x]$ denotes the integral part of x.

† King$^{(1)}$ states without qualification that his equation (14), corresponding to (12) here, represents a maximum. This is, of course, true for his experimental values.

or, assuming z_1 and s_2 to be the distances of the first nodes from either side of the dielectric,

$$s_2=z_1.$$

Similarly from (11) and (13) $s'+s_2=(\nu+\frac{1}{2})\lambda$; in particular $s_2=z_1$.

These results show that for maximum and minimum shift the dielectric slab is so placed that nodes on both sides are symmetrical with respect to the slab.

(b) *Boundary Effects in a System obeying the Resonance Condition.*

Using equations (1), (5), (6), (7), the following expressions are obtained for E_0, E_1, and E_2:

$$E_0=-2A_0ie^{ipt}\sin\beta z,\qquad 0\leqslant z\leqslant z_1, \quad \ldots\ldots \quad (14)$$

$$\begin{aligned}E_1&=-2A_0ie^{ipt}\left(\cos\beta_1 s\sin\beta z_1+\frac{1}{n}\sin\beta_1 s\cos\beta z_1\right)\\&=-2A_0ie^{ipt}\sqrt{\sin^2\beta z_1+\frac{1}{n^2}\cos^2\beta z_1}\,.\sin\{\beta_1 s+\tan^{-1}(n\tan\beta z_1)\}, \quad . \quad (15)\\&\qquad\qquad 0\leqslant s\leqslant z_2-z_1,\end{aligned}$$

where

$$s=z-z_1,$$

$$\begin{aligned}E_2&=-2A_0ie^{ipt}\frac{1}{n}\{\sin\beta s''(n\cos\beta z_1\cos\beta_1 s_1-n^2\sin\beta z_1\sin\beta_1 s_1)\\&\qquad+\cos\beta s''(\cos\beta z_1\sin\beta_1 s_1+n\sin\beta z_1\cos\beta_1 s_1)\}\\&=-2A_0ie^{ipt}R\sin(\beta s''+\phi),\qquad 0\leqslant s''\leqslant z_3-z_2, \quad \ldots\ldots \quad (16)\end{aligned}$$

where

$$R=\left\{(\cos\beta z_1\cos\beta_1 s_1-n\sin\beta z_1\sin\beta_1 s_1)^2+\left(\frac{1}{n}\cos\beta z_1\sin\beta_1 s_1+\sin\beta z_1\cos\beta_1 s_1\right)^2\right\}^{1/2}$$

$$\tan\phi=\frac{\frac{1}{n}\tan\beta_1 s_1+\tan\beta z_1}{1-n\tan\beta_1 s_1\tan\beta z_1}$$

and

$$s''=z-z_2.$$

Before reduction E_1 and E_2 are observed to consist of four waves, two travelling in the z direction and two in the $-z$ direction. The expressions given are the stationary waves which are the resultants of these components.

At $\quad z=z_1 \quad E_0=-2A_0ie^{ipt}\sin\beta z_1$,

at $\quad s=0 \quad E_1=-2A_0ie^{ipt}\sin\beta z_1$.

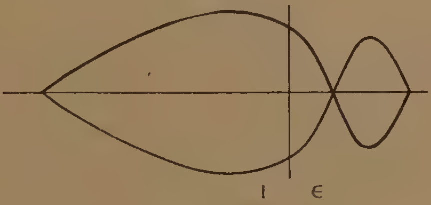
At the boundary therefore E_0 and E_1 have the same intensity, but the phases are not equal, being respectively βz_1 and $\tan^{-1}(n \tan \beta z_1)$. It is easy to show, remembering the relation $\beta_1 = \sqrt{\epsilon}\beta$, that

$$\left(\frac{\partial E_0}{\partial z}\right)_{z=z_1} = \left(\frac{\partial E_1}{\partial s}\right)_{s=0} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Similar results can be shown to hold for the second boundary.

Thus, although for any progressive train of waves crossing a boundary of dielectrics there will be a change of intensity by reflexion and no change of phase between incident and transmitted beam, in this case for the condition of resonance the reflexions at the bridge and at dielectric boundaries are such that the resultant stationary wave shows no change of intensity at the boundary *, but does show a phase change. From the above results it is seen that the magnitude of the phase change is dependent upon z_1, *i. e.*, upon the position of the dielectric relative to the original standing wave system. In this lies the explanation of the variation of bridge shift with dielectric position. From (17) we see that the phase change at a boundary is always such as to make the gradient of electric intensity continuous. Fig. 3 illustrates these conditions; the curves represent extreme values of E.

Fig. 3.



The expressions for the magnetic field, viz.,

$$\left.\begin{aligned}
H_0 &= 2A_0 e^{ipt} \cos \beta z, \\
H_1 &= 2A_0 e^{ipt} \sqrt{\epsilon} \sqrt{\sin^2 \beta z_1 + \frac{1}{n^2} \cos^2 \beta z_1} \,.\, \cos\{\beta_1 s + \tan^{-1}(n \tan \beta z_1)\}, \\
H_2 &= 2A_0 e^{ipt} R \cos(\beta s'' + \phi),
\end{aligned}\right\} \quad . \quad (18)$$

are similar to those of the electric field displaced by $\pi/2$, with the exception that a factor $\sqrt{\epsilon}$ appears in H_1 (this factor is present, equal to unity, in H_0 and H_2). The gradient of H is now discontinuous at a boundary (except when zero). The relative amplitudes of E and H follow the usual law for propagation of plane waves, viz., $H = \sqrt{\epsilon} E$, and the discontinuity of gradient is such that loops of H are coincident with nodes of E.

---

* This does not mean that the amplitudes of E_0, E_1, E_2 are equal.

As the potential between the wires is obtained by a line integration of the complete expression for the electric force, potential distribution is the same as distribution of E ; current in the wires, being an integral of magnetic force, follows the distribution of H.

A decrease δ in the resonant length of the line is due to a phase increase of $\beta\delta$ in both E and H. This total phase increase will be due to :

2 phase shifts at dielectric boundaries+phase increase due to optical path difference between dielectric and the same thickness of air.

This can be expressed as

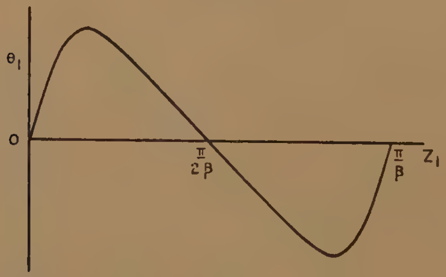
$$\beta\delta=\theta_1+\theta_2+(\beta_1-\beta)s_1.$$

From (14), (15), (16),

$$\theta_1=\tan^{-1}(n\tan\beta z_1)-\beta z_1, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad (19)$$

$$\theta_2=\tan^{-1}\left\{\frac{\frac{1}{n}\tan\beta_1 s_1+\tan\beta z_1}{1-n\tan\beta z_1\tan\beta_1 s_1}\right\}-\tan^{-1}(n\tan\beta z_1)-\beta_1 s_1. \qquad (20)$$

Fig. 4.



The relation between θ_1 and z_1, for $n=4$, is shown graphically in fig. 4. From it we see that if $0<z_1<\frac{1}{4}\lambda$ there will be an advance of phase at the boundary, and a retardation for $\frac{1}{4}\lambda<z_1<\frac{1}{2}\lambda$. Turning points on the curve are given by

$$\tan\beta z_1=\pm\frac{1}{\sqrt{n}},$$

and so the skewness increases with n.

If we replace $\tan\beta z_1$ by its value in terms of s_1 and s_2 given by (8) we obtain

$$\theta_2=\tan^{-1}(n\tan\beta s_2)-\beta s_2,$$

which is of the same form as θ_1. For the conditions of $\delta_{\max}$ and $\delta_{\min}$

$$\theta_1=\theta_2,$$

and since also $z_1 = s_2$ for these cases, we see from fig. 4 that,

if $$0 < z_1 = s_2 < \tfrac{1}{4}\lambda,$$

δ will be a maximum ;

if $$\tfrac{1}{4}\lambda < z_1 = s_2 < \tfrac{1}{2}\lambda,$$

δ will be a minimum.

The phase change due to the optical path difference is, for $\delta_{\max}$, increased by two equal boundary shifts, and, for $\delta_{\min}$, decreased by two equal boundary shifts.

(c) *Relative Amplitudes.*

$$|E_0| = 2A_0,$$

$$|E_1| = 2A_0\sqrt{\sin^2 \beta z_1 + \frac{1}{n^2}\cos^2 \beta z_1},$$

$$|E_2| = 2A_0\left\{(\cos \beta z_1 \cos \beta_1 s_1 - n \sin \beta z_1 \sin \beta_1 s_1)^2 + \left(\frac{1}{n}\cos \beta z_1 \sin \beta_1 s_1 + \sin \beta z_1 \cos \beta_1 s_1\right)^2\right\}^{\frac{1}{2}}.$$

It can be seen that $|E_1|$ varies between the limits $|E_0|$ and $\frac{1}{\sqrt{\epsilon}}|E_0|$, and $|E_2|$ between limits $\sqrt{\epsilon}|E_0|$ and $\frac{1}{\sqrt{\epsilon}}|E_0|$. This is also the potential distribution. The limits for the magnetic field, and hence current, are

$$|H_0| \leqslant |H_1| \leqslant \sqrt{\epsilon}|H_0|,$$

$$\frac{1}{\sqrt{\epsilon}}|H_0| \leqslant |H_2| \leqslant \sqrt{\epsilon}|H_0|.$$

(d) *Summary.*

It has been shown that, on surrounding part of the length of a Lecher wire system by a dielectric, a decrease in the resonant length is produced, which is dependent on the position of the dielectric slab. For a system in resonance the following are true :

(1) At a dielectric boundary the electric and magnetic intensities are continuous.

(2) A phase change is produced the same in both electric and magnetic components and such that the gradient of electric intensity is continuous. Loops of magnetic force coincide with nodes of electric force.

(3) The amplitudes of the stationary waves in the three zones are in general different and vary relatively between definite limits. The typical general case of fig. 5 illustrates these results, solid lines being of electric intensity and dotted lines of magnetic intensity.

(4) For maximum and minimum bridge displacement the wave distribution is symmetrical with respect to the dielectric slab. Examples of this are shown in fig. 6.

Fig. 5.

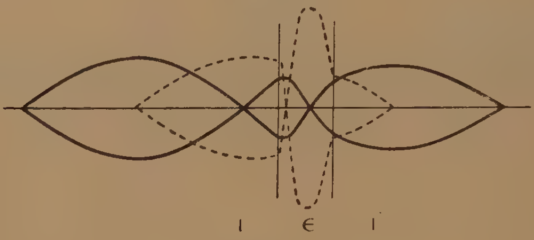
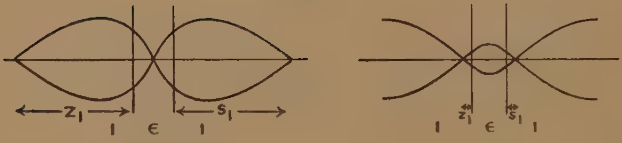


Fig. 6.



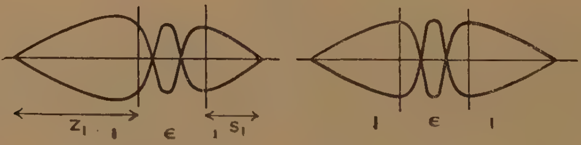
MINIMUM

MAXIMUM

(e) *Particular Cases.*

An interesting particular case, fig. 7 (*a*), is where s_1 is a whole number of dielectric half-wave-lengths. Then $\theta_1+\theta_2=0$ and δ depends only on the optical path difference, and is therefore constant for all positions of the slab. In this case the part s_1 of the wave is the complement of the part z_1.

Fig. 7.



(*a*) (*b*)

When a node or a loop coincides with a dielectric boundary the phase change there is zero. When a loop coincides with the boundary the amplitudes on both sides of it are equal. Fig. 7 (*b*) illustrates this as a particular case of 7 (*a*); all three amplitudes are here equal.

Previous Investigations.

A number of investigators have employed experimental arrangements similar to that considered here, without fully appreciating the principles involved.

Smith-Rose and McPetrie (5), in an investigation on the dielectric constant of soil, used a box of soil with two parallel wires passing through it. One end was made to coincide with a voltage node, "so that the effect of reflexion at the boundary between soil and air would be reduced to a minimum"; the other end, however, did not coincide with a node or loop (except perhaps fortuitously). The formula stated was the result of equating bridge shift to optical path difference. Neglect of the second boundary effect invalidates this formula and the results obtained.

The results of Banerjee and Joshi (6), who used the same formula in a similar investigation, are also in error.

Seeberger (7) was puzzled by the variation of bridge shift with dielectric position, and dismissed it as "unübersichtlich."

Hormell (8) alone seems to have appreciated the boundary effect, although he does not say so explicitly. In a measurement of the dielectric constant of paraffin wax he cut the slab until each end coincided with a current loop (thus making θ_1 and θ_2 zero).

PART II.

Introduction.

Recently there has arisen much interest in the properties of hollow metal tubes as guides for micro-waves *. Three papers on the subject have been quoted (2), (3), (4). A later paper by Schelkunoff (9) gives a complete bibliography; a list of papers subsequent to that is appended here.

Briefly electromagnetic waves may be propagated through a single hollow tube of any section, subject to certain conditions. A series of characteristic wave-forms is possible, there being a limiting frequency for each type below which no propagation occurs. For circular tubes in particular the free space wave-length corresponding to a limiting, or critical, frequency for a tube of radius a is given by $\lambda_c=\frac{2\pi\sqrt{\epsilon}}{k}$, where ϵ is the dielectric constant of the material filling the tube and k has a value satisfying the equation

$$J_\nu(ka)=0$$

or the equation

$$J_\nu'(ka)=0, \quad \left(i.\,e.,\ \frac{d}{d(ka)}J_\nu(ka)=0\right).$$

$J_\nu(ka)$ is the Bessel function of the first kind of integral order ν.

---

* No suitable neat expression has been evolved in English for these transmission lines. Southworth's term "wave guide" is not sufficiently particular, as any system of conductors or non-conductors is a wave guide, but it has come into use in the restricted sense employed here.

The largest values of λ for the respective equations are

$$\lambda=2{\cdot}61\, a\sqrt{\epsilon}$$

and

$$\lambda'=3{\cdot}41\, a\sqrt{\epsilon},$$

thus limiting practical applications to micro-waves.

The wave-length inside the tube is not $\dfrac{\lambda}{\sqrt{\epsilon}}$ as for a parallel wire system, but is given for the ideal non-dissipative case by the equation

$$\frac{1}{\lambda_{\text{tube}}}=\sqrt{\frac{\epsilon}{\lambda^2}-\left(\frac{k}{2\pi}\right)^2}.$$

Also the wave inside the tube is not a plane wave; there exists always a component of electric intensity or of magnetic intensity or of both in the direction of propagation. Waves are broadly classifiable as types E and H. E-waves are those depending on solutions of $J_\nu(ka)=0$ and have zero axial magnetic force; H-waves depend on solutions of $J_\nu'(ka)=0$ and have zero axial electric force.

Theory.

(a) *H-waves.*

The expressions for E and H for an H-wave in a wave guide are given in cylindrical coordinates ρ, ϕ, z by

$$H_z=J_\nu(k\rho)\cos\nu\phi(Ae^{i(pt-\beta z)}+Be^{i(pt+\beta z)}),$$

$$E_z=0,$$

$$E_\rho=\frac{\nu p}{ik^2\rho c}J_\nu(k\rho)\sin\nu\phi(Ae^{i(pt-\beta z)}+Be^{i(pt+\beta z)}),$$

$$E_\phi=-\frac{p}{ikc}J_\nu'(k\rho)\cos\nu\phi(Ae^{i(pt-\beta z)}+Be^{i(pt+\beta z)}),$$

$$H_\rho=\frac{\beta}{ik}J_\nu'(k\rho)\cos\nu\phi(Ae^{i(pt-\beta z)}-Be^{i(pt+\beta z)}),$$

$$H_\phi=-\frac{\nu\beta}{ik^2\rho}J_\nu(k\rho)\sin\nu\phi(Ae^{i(pt-\beta z)}-Be^{i(pt+\beta z)}).$$

k in the above equations has a value given by

$$J_\nu'(ka)=0$$

and

$$\beta\equiv\frac{2\pi}{\lambda_{\text{tube}}}=2\pi\sqrt{\frac{\epsilon}{\lambda^2}-\left(\frac{k}{2\pi}\right)^2}.$$

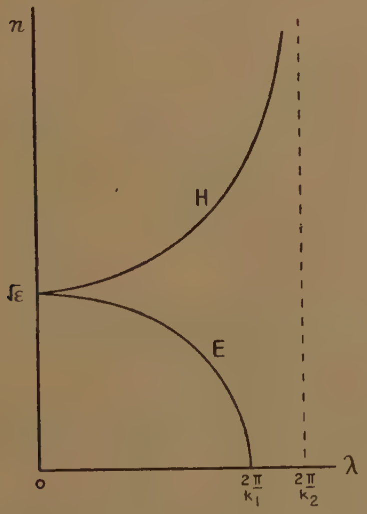
Let us take a tube, part of which between two diametral planes is filled with dielectric, the rest being filled with air, and let us define zones 0, 1, 2 as in Part I. The air-dielectric boundary conditions are continuity of the tangential components E_ρ, E_ϕ, H_ρ, H_ϕ; for H_z, which is perpendicular to the boundary plane, the condition is continuity of magnetic induction, but since permeability is assumed unity in both air and dielectric this reduces to continuity of H_z. The application of these conditions gives for H_z, E_ρ, E_ϕ

$$A_0 e^{-i\beta z_1}+B_0 e^{i\beta z_1}=A_1 e^{-i\beta_1 z_1}+B_1 e^{i\beta_1 z_1},$$
$$A_1 e^{-i\beta_1 z_2}+B_1 e^{i\beta_1 z_2}=A_2 e^{-i\beta z_2}+B_2 e^{i\beta z_2},$$

and for H_ρ, H_ϕ

$$\beta(A_0 e^{-i\beta z_1}-B_0 e^{i\beta z_1})=\beta_1(A_1 e^{-i\beta_1 z_1}-B_1 e^{i\beta_1 z_1}),$$
$$\beta_1(A_1 e^{-i\beta_1 z_2}-B_1 e^{i\beta_1 z_2})=\beta(A_2 e^{-i\beta z_2}-B_2 e^{i\beta z_2}).$$

Fig. 8.



These are formally identical with equations (2); n has the value β_1/β, and hence now

$$n=\sqrt{\frac{\epsilon-\left(\frac{k\lambda}{2\pi}\right)^2}{1-\left(\frac{k\lambda}{2\pi}\right)^2}} \qquad (cf.\ (3)) \quad . \ . \ . \ . \ . \quad (21)$$

If the end of the tube is closed by a plane reflector, (4) will express this terminal condition. Now choose the origin of coordinates such that (5) is also satisfied. Since equations (2), (4), (5) all hold for the wave guide, the fundamental equation

$$n \cot \beta_1 s_1(\cot \beta s'-\cot \beta s_2)+\cot \beta s' \cot \beta s_2+n^2=0 \quad . \ . \ . \quad (9)$$

will also hold. It will give a shift curve with maxima and minima given as before by (10) to (13). Since n, β, and β_1 have assumed new values the consequences of (9) may differ from those obtained in Part I.

An obvious difference is that, whereas in Part I. n was constant equal to $\sqrt{\epsilon}$, now n given by (21) is a function of ϵ and of λ. Since to allow of transmission both numerator and denominator of (21) must be real, λ is restricted to values $<\frac{2\pi}{k}$. The portion marked H in fig. 8 shows the mode of variation of n with λ in the permissible range. It increases from a value $\sqrt{\epsilon}$ for infinitely high frequencies to an infinite value at the cut-off point.

It is obvious from comparison with the expressions for the components of E and H in Part I. that the following results will hold (omitting non-significant factors):—

$$H_{z0}\sim E_{\rho 0}\sim E_{\phi 0}\sim \sin \beta z,$$

$$H_{\rho 0}\sim H_{\phi 0}\sim \beta \cos \beta z,$$

$$H_{z1}\sim E_{\rho 1}\sim E_{\phi 1}\sim \sqrt{\sin^2 \beta z_1+\frac{1}{n^2}\cos^2 \beta z_1}\,.\sin\{\beta_1 s+\tan^{-1}(n\tan\beta z_1)\},$$

$$H_{\rho 1}\sim H_{\phi 1}\sim \beta_1\sqrt{\sin^2 \beta z_1+\frac{1}{n^2}\cos^2 \beta z_1}\,.\cos\{\beta_1 s+\tan^{-1}(n\tan\beta z_1)\},$$

etc.

Once again, in addition to continuity of intensity at a boundary, we have an adjustment of relative amplitudes such that the electric field gradient (and also the gradient of H_z) is continuous.

With the appropriate changes the results of Part I. are applicable, and similar distribution diagrams can be drawn.

(b) *E-waves.*

For the case of a wave whose magnetic vector is wholly transverse the propagation equations in the wave guide are

$$E_z=J_\nu(k\rho)\cos\nu\phi(A'e^{i(pt-\beta z)}-B'e^{i(pt+\beta z)}),$$

$$H_z=0,$$

$$E_\rho=\frac{\beta}{ik}J_\nu'(k\rho)\cos\nu\phi(A'e^{i(pt-\beta z)}+B'e^{i(pt+\beta z)}),$$

$$E_\phi=-\frac{\nu\beta}{ik^2\rho}J_\nu(k\rho)\sin\nu\phi(A'e^{i(pt-\beta z)}+B'e^{i(pt+\beta z)}),$$

$$H_\rho=\frac{p\nu\epsilon}{ik^2\rho c}J_\nu(k\rho)\sin\nu\phi(A'e^{i(pt-\beta z)}-B'e^{i(pt+\beta z)}),$$

$$H_\phi=\frac{p\epsilon}{ikc}J_\nu'(k\rho)\cos\nu\phi(A'e^{i(pt-\beta z)}-B'e^{i(pt+\beta z)}).$$

k in these equations has a value given by $J_\nu(ka)=0$, and the propagation constant

$$\beta \equiv \frac{2\pi}{\lambda_{\text{tube}}} = 2\pi\sqrt{\frac{\epsilon}{\lambda^2}-\left(\frac{k}{2\pi}\right)^2}.$$

For the interior of the tube divided into zones as before the air-dielectric boundary conditions are continuity of the tangential components; for E_z the condition is continuity of displacement. The application of these conditions gives for E_ρ and E_ϕ

$$\beta(A_0' e^{-i\beta z_1}+B_0' e^{i\beta z_1})=\beta_1(A_1' e^{-i\beta_1 z_1}+B_1' e^{i\beta_1 z_1}),$$
$$\beta_1(A_1' e^{-i\beta_1 z_2}+B_1' e^{i\beta_1 z_2})=\beta(A_2' e^{-i\beta z_2}+B_2' e^{i\beta z_2}),$$

and for E_z, H_ρ, and H_ϕ

$$A_0' e^{-i\beta z_1}-B_0' e^{i\beta z_1}=\epsilon(A_1' e^{-i\beta_1 z_1}-B_1' e^{i\beta_1 z_1}),$$
$$\epsilon(A_1' e^{-i\beta_1 z_2}-B_1' e^{i\beta_1 z_2})=A_2' e^{-i\beta z_2}-B_2' e^{i\beta z_2}.$$

If in these equations we put

$$\beta A_0'=A_0, \quad \beta_1 A_1'=A_1, \quad \beta A_2'=A_2,$$

and similar values for the B's, they reduce to equations (2) formally. This time n has the value $\frac{\epsilon\beta}{\beta_1}$, and hence

$$n=\epsilon\sqrt{\frac{1-\left(\frac{k\lambda}{2\pi}\right)^2}{\epsilon-\left(\frac{k\lambda}{2\pi}\right)^2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Applying as before the terminal and origin conditions (4) and (5), we get again the resonance relation (9) *.

* Schelkunoff (9) has shown that the differential equations of wave motion for E-waves can be expressed as

$$\frac{\partial V}{\partial z}=\left(\frac{ip}{c}+\frac{k^2c}{ip\epsilon}\right)A=-ZA,$$
$$\frac{\partial A}{\partial z}=-\frac{ip\epsilon}{c}V=-YV,$$

where V is a scalar potential function such that the transverse electric field $E_t=-\text{grad}_t V$, and A is a vector parallel to the z-axis such that H=curl A and proportional to the longitudinal displacement current. These correspond exactly to the differential equations for current and potential in parallel wire or coaxial transmission lines, Z being the distributed series impedance and Y the distributed shunt admittance. We can therefore define the parameter $\sqrt{\frac{Z}{Y}}$ as the "characteristic impedance" of the tube. King (1) defines n in (9) as the ratio of the characteristic impedances of a parallel wire system in air and

Again n is a function of λ varying this time in the range $\sqrt{\epsilon}$ to 0, as shown in the curve marked E of fig. 8.

In particular, for any value of ϵ by appropriate choice of λ, n can be made unity. Substituting this value in (9), and reducing, we find

$$\begin{aligned}\cot\beta_1 s_1 &= \cot\beta(s'-s_2)\\ &= \cot\beta(s_1+\delta)\end{aligned}$$

or

$$\beta_1 s_1=\beta(s_1+\delta). \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (23)$$

Thus the shift curve of fig. 2 has degenerated to a straight line, the shift being now constant and independent of the dielectric position. The value of λ for which $n=1$ is given by

$$\lambda=\frac{2\pi}{k}\sqrt{\frac{\epsilon}{\epsilon+1}}=\lambda_c\sqrt{\frac{\epsilon}{\epsilon+1}}, \quad . \; . \; . \; . \; . \; . \; . \quad (24)$$

where λ_c is the critical wave-length for the particular wave-type employed. If this point is found an easy means of determining ϵ is thus available.

For values of $n>1$ the conditions for $\delta_{\max}$ and $\delta_{\min}$ will hold as in Part I., but if $n<1$ these conditions will be reversed, *e. g.*, for $\left[\frac{2s_1}{\lambda_1}\right]$ odd

if $n>1$, $\delta_{\max}$ is given by (12);

if $n<1$, $\delta_{\max}$ is given by (13).

If we let δ_1 be the value of δ given by (13), and δ_2 the value given by (12), then

$$\delta_1-\delta_2=\frac{2}{\beta}\tan^{-1}\left\{\frac{1}{2}\left(n-\frac{1}{n}\right)\sin\beta_1 s_1\right\}. \quad . \; . \; . \; . \; . \quad (25)$$

---

in the dielectric. Adapting this definition of n, and using the above expression for the characteristic impedance, we get

$$n=\sqrt{\frac{\dfrac{ip}{c}+\dfrac{k^2c}{ip}}{\dfrac{ip}{c}}\cdot\frac{\dfrac{ip\epsilon}{c}}{\dfrac{ip}{c}+\dfrac{k^2c}{ip\epsilon}}},$$

which reduces to

$$\epsilon\sqrt{\frac{1-\left(\dfrac{k\lambda}{2\pi}\right)^2}{\epsilon-\left(\dfrac{k\lambda}{2\pi}\right)^2}},$$

the expression found in (22).

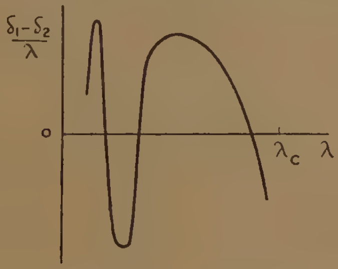
For H-waves equation (21) can be similarly obtained by using Schelkunoff's values of

$$Z=\frac{ip}{c}, \quad Y=\frac{ip\epsilon}{c}+\frac{k^2c}{ip}.$$

Fig. 9 shows the variation of this expression with λ, $\delta_1-\delta_2$ being expressed in units of wave-length. The zero nearest to λ_c is the point where $n=1$; the other zeros are points where $\sin\beta_1 s_1=0$, and the variation of this term produces a rapid alternation as $\lambda\to 0$. The value of λ for the first zero is independent of the thickness of the dielectric, the others are dependent on it.

Expressions similar to those for H-waves can be obtained for the various field components in each zone. The electric vector in this case is not confined to one plane, the lines of electric force being three dimensional twisted curves whose transverse and longitudinal components obey different boundary laws. This being so the relations of gradient of E at a boundary are not so simple as for the parallel wire system and for H-waves in the wave guide. The significant parts of the relations are

Fig. 9.



$$E_{z0}\sim\cos\beta z,$$
$$E_{\rho 0}\sim E_{\phi 0}\sim\beta\sin\beta z,$$
$$H_{\rho 0}\sim H_{\phi 0}\sim\sin\beta z,$$
$$E_{z1}\sim\frac{\beta}{\beta_1}R'\cos(\beta_1 s+\phi'),$$
$$E_{\rho 1}\sim E_{\phi 1}\sim\frac{\beta}{\beta_1}\beta_1 R'\sin(\beta_1 s+\phi'),$$
$$H_{\rho 1}\sim H_{\phi 1}\sim\frac{\beta}{\beta_1}\epsilon R'\sin(\beta_1 s+\phi'),$$

where
$$R'=\sqrt{\sin^2\beta z_1+\frac{1}{n^2}\cos^2\beta z_1}$$

and
$$\tan\phi'=n\tan\beta z_1,\quad \text{etc.}$$

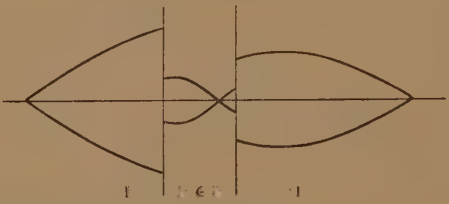
All the transverse components are continuous at the boundary; for the longitudinal component there is continuity of displacement. The ratio of boundary gradients for E_ρ and E_ϕ is $\beta_1^2/\beta^2\epsilon$, which is not equal to unity. Thus the simple diagrammatic method of tracing the wave distribution is not now available. However, the gradient of E_z is continuous at the boundary although actual intensities are in the ratio of ϵ. These facts enable us to draw rough distribution diagrams such as fig. 10, which represents E_z.

The relative simplicity of the components for an H-wave is due to the fact that we have the permeability terms unity. Had this not been so the introduction of a term μ for the dielectric would have made the equations for H-waves similar to those for E-waves with ϵ replaced by μ.

(c) *Summary.*

It has been shown that the decrease in resonant length caused by filling part of a hollow tube wave guide with dielectric material is a function of the position of this dielectric cylinder. The form of the function is essentially similar to that for the parallel wire system, but has important differences.

Fig. 10.



The E and H types of wave are studied.

For the H type the distribution is similar to that described in Part I.

The E type is more complex and is not readily illustrated graphically. It has an important particular case where for a definite wave-length the shift remains constant for all positions of the dielectric.

Experimental Work.

The results here obtained have been adequately verified for E- and H-waves in a hollow tube wave guide. A description of these experiments will be published shortly, together with an account of the use of the theory for determination of high-frequency dielectric constants.

Acknowledgments.

The work was done in the laboratories of the Natural Philosophy Department of the University of Glasgow, and owes much to the constant advice of Dr. John Thomson, whose kindness the author gratefully acknowledges. Thanks are also due to Professor Taylor Jones for facilities for the work and encouragement in it.

Bibliography.

Papers on transmission of electromagnetic waves through hollow metal tubes, subsequent to Schelkunoff(9).

G. C. Southworth, "Some Fundamental Experiments with Wave Guides," Proc. Inst. Radio Engrs. xxv. p. 807 (1937).

L. Page and N. I. Adams, Jr., "Electromagnetic Waves in Conducting Tubes," Phys. Rev. lii. p. 647 (1937).

G. C. Southworth, "New Experimental Methods applicable to Ultra-short Waves," J. Appl. Phys. viii. p. 660 (1937).

J. Saphores, "General Properties of Dielectric Guides," Elect. Commun. xvi. p. 346 (1938).

L. Brillouin, "Theoretical Study of Dielectric Cables," Elect. Commun. xvi. p. 350 (1938).

A. G. Clavier, "Théorie des câbles diélectriques cylindriques et relations avec la théorie des câbles coaxiaux," *Bull. Soc. Franç. Élect.* viii. p. 355 (1938).

E. Kašpar, "Experimentelle Untersuchung der elektromagnetischen Wellen an dielektrischen Drähten," *Ann. Phys., Lpz.* xxxii. p. 353 (1938).

S. A. Schelkunoff, "Note on certain Guided Waves in slightly Noncircular Tubes," J. Appl. Phys. ix. p. 484 (1938).

H. W. Droste, "Ultrahochfrequenz-Übertragang längs zylindrischen Leitern und Nichtleitern," *Telegr.- u. FernsprTech.* xxvii. pp. 199, 273, 310, 337 (1938).

L. J. Chu, "Electromagnetic Waves in Elliptic Hollow Pipes of Metal," J. Appl. Phys. ix. p. 583 (1938).

A. G. Clavier and V. Altovsky, "Étude complémentaire sur les coefficients d'atténuation dans les câbles diélectriques cylindriques et les câbles coaxiaux," *Bull. Soc. Franç. Élect.* viii. p. 793 (1938).

H. E. Hartig and C. E. Swanson, "'Transverse' Acoustic Waves in Rigid Tubes," Phys. Rev. liv. p. 618 (1938).

H. E. Hartig and A. W. Mellon, "The Transmission of Damped Electromagnetic Waves through small Hollow Metal Tubes," Phys. Rev. liv. p. 646 (1938).

L. Brillouin, "Propagation des ondes sur les câbles diélectriques," *Bull. Soc. Franç. Élect.* viii. p. 899 (1938).

H. Buchholz, "Die Quasioptik der Ultrakurzwellenleiter," *Elekt. Nachr.-Tech.* xv. p. 297 (1938).

L. J. Chu and W. L. Barrow, "Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section," Proc. Inst. Radio Engrs. xxvi. p. 1520 (1938).

W. L. Barrow and L. J. Chu, "Theory of the Electromagnetic Horn," Proc. Inst. Radio Engrs. xxvii. p. 51 (1939).

A. G. Clavier, "Theoretical Relationships of Dielectric Guides (Cylindrical) and Co-axial Cables," Elect. Commun. xvii. p. 276 (1939).

H. Buchholz, "Die Einfluss der Krümmung von recteckigen Hohlleitern auf das Phasenmass ultrakurzer Wellen," *Elekt. Nachr.-Tech.* xvi. p. 73 (1939).

H. Riedel, "Der rechteckförmige metallische Hohlleiter als Übertragungsweg elektromagnetischer Wellen," *Hochfrequenztech. u. Elektroakust.* liii. p. 122 (1939).

A. Klemt, "Beitrag zur Aus–reitung elektromagnetischer Wellen an dielektrischen Drähten," *Funktech. Mh.* p. 122 (1939).

References.

(1) R. King, Rev. Sci. Instrum. viii. p. 201 (1937).
(2) W. L. Barrow, Proc. Inst. Radio Engrs. xxiv. p. 1298 (1936).
(3) G. C. Southworth, Bell Syst. Tech. J. xv. p. 284 (1936).
(4) J. R. Carson, S. P. Mead, and S. A. Schelkunoff, Bell Syst. Tech. J. xv. p. 310 (1936).
(5) R. L. Smith-Rose and J. S. McPetrie, Proc. Phys. Soc. xlvi. p. 649 (1934).
(6) S. S. Banerjee and R. D. Joshi, Phil. Mag. xxv. p. 1025 (1938).
(7) M. Seeberger, *Ann. Phys., Lpz.* xvi. p. 77 (1933).
(8) W. G. Hormell, Phil. Mag. iii. p. 52 (1902).
(9) S. A. Schelkunoff, Proc. Inst. Radio Engrs. xxv. p. 1457 (1937).

May 1939.

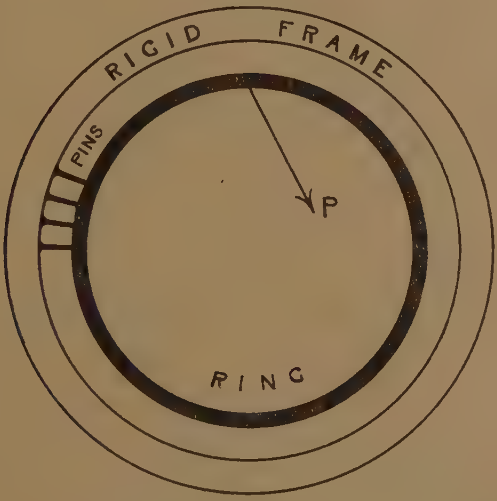
LIV. *The Stresses in a Circular Ring supported on Frictionless Radial Pins.*

By Professor A. J. Sutton Pippard, M.B.E., D.Sc.,
Imperial College, London *.

[Received August 10, 1939.]

The problem considered occurred recently in design, and is shown diagrammatically in fig. 1.

Fig. 1.



A circular ring had to be supported on a stiff frame made of material having a very different coefficient of expansion from that of the ring.

It was essential to avoid distortion of the ring which would arise from temperature changes if the two were firmly joined, and connexion between them was therefore made by a large number of radial pins rigidly attached to the outer frame and sliding freely in holes in the ring. These pins offer no restraint to radial movements, but their flexural stiffness resists tangential displacements.

The ring is strained by a load applied to any point on its circumference, and it is convenient for purposes of analysis to resolve this load into its radial and tangential components and to treat these separately. The net result is then obtained by superposing the two separate solutions.

---

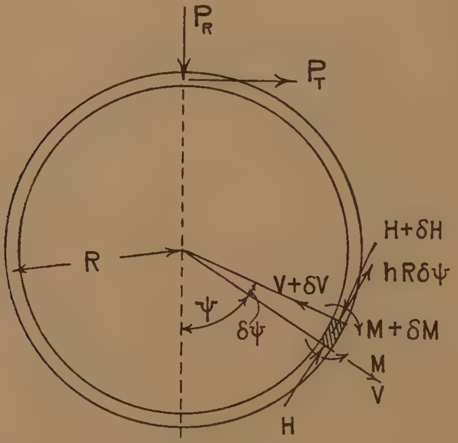
* Communicated by the Author.

The problem is similar in many respects to that of the wheel with non-radial spokes, a solution for which has been given previously *, and the same method is used to obtain a solution.

The pins are sufficiently numerous for their effect to be replaced by a hypothetical continuous annular sheet covering the ring, the properties of this sheet being that while it has no radial stiffness it can resist tangential forces.

In fig. 2 the load P is shown resolved into its components P_R and P_T. The origin of coordinates is at the point on the ring diametrically opposite to the point of application of the load.

Fig. 2.



Let R be the radius to the centroid of the ring,
EI, the flexural rigidity of the cross-section,
a, the cross-sectional area,
h, the intensity of tangential force exerted by the sheet on the ring at an angular distance ψ from the origin,
u, the outward radial displacement at this point,
and v, the anti-clockwise tangential displacement.

The resultant actions in the ring at this point are a tangential force H, a radial force V, and a bending moment M, the positive directions of these being as shown in fig. 2.

---

* "The Stresses in a Wire Wheel with Non-radial Spokes under Loads applied to the Rim," A. J. Sutton Pippard and Miss M. J. White, Phil. Mag. xiv. p. 209 (August 1932).

Then
$$H=-\frac{Ea}{R}\left(u+\frac{\partial v}{\partial\psi}\right). \qquad (1)$$

and
$$M=\frac{EI}{R^2}\left(u+\frac{\partial^2 u}{\partial\psi^2}\right). \qquad (2)$$

By resolution of the forces acting on a small element of the ring subtending an angle $\delta\psi$ at the centre the following equations are obtained :—

$$-V+hR=\frac{\partial H}{\partial\psi}, \qquad (3)$$

$$\frac{\partial V}{\partial\psi}=H, \qquad (4)$$

$$\frac{\partial M}{\partial\psi}=VR. \qquad (5)$$

Equations (2) to (4) lead to

$$\frac{\partial^5 u}{\partial\psi^5}+2\frac{\partial^3 u}{\partial\psi^3}+\frac{\partial u}{\partial\psi}-\frac{R^4 h}{EI}=0. \qquad (6)$$

Let K be the force required per unit length of the perimeter of the ring to produce a unit tangential displacement ;

so that
$$h=-Kv.$$

When there are N radial pins, each of which requires a force W to produce a unit deflexion,

$$K=\frac{NW}{2\pi R}.$$

The condition that the ring shall undergo no change of length or, alternatively, that the strain energy due to tangential tension or compression is negligible compared with that due to bending, is

$$u=-\frac{\partial v}{\partial\psi}. \qquad (7)$$

Hence
$$\frac{\partial h}{\partial\psi}=Ku.$$

On differentiating (6) and substituting this value the result is obtained,

$$\frac{\partial^6 u}{\partial\psi^6}+2\frac{\partial^4 u}{\partial\psi^4}+\frac{\partial^2 u}{\partial\psi^2}-Qu=0, \qquad (8)$$

where
$$Q=\frac{R^4K}{EI}.$$

The complete solution of this equation is

$$u=\text{A}\sinh\gamma\psi+\text{B}\cos\alpha\psi\sinh\beta\psi+\text{C}\sin\alpha\psi\cosh\beta\psi$$
$$+\text{D}\cosh\gamma\psi+\text{E}\cos\alpha\psi\cosh\beta\psi+\text{F}\sin\alpha\psi\sinh\beta\psi, \quad . \quad . \quad . \quad (9)$$

where A, B, C, D, E and F are constants of integration to be determined.

The values of α, β, and γ are obtained from the following.

Let $$\text{M}=\frac{1}{27}+\frac{\text{Q}}{2}$$

and $$\text{N}=\sqrt{\frac{\text{Q}^2}{4}+\frac{\text{Q}}{27}}.$$

Then if $$\text{X}=(\text{M}+\text{N})^{\frac{1}{3}}+(\text{M}-\text{N})^{\frac{1}{3}},$$
$$\alpha=\tfrac{1}{2}\sqrt{3\text{X}+2},$$
$$\beta=\frac{1}{2\sqrt{3}}\sqrt{3\text{X}-2},$$

and $$\gamma=2\beta.$$

CASE 1.—*Radial Load at* $\psi=\pi$.

From considerations of symmetry it is clear that the values of u for equal positive and negative values of ψ are equal, and so the constants A, B, and C in (9) are all zero, and for this case

$$u=\text{D}\cosh 2\beta\psi+\text{E}\cos\alpha\psi\cosh\beta\psi+\text{F}\sin\alpha\psi\sinh\beta\psi \quad . \quad . \quad . \quad . \quad (10)$$

and

$$v=-\left[\frac{\text{D}}{2\beta}\sinh 2\beta\psi+\frac{\text{E}\beta-\text{F}\alpha}{\alpha^2+\beta^2}\cos\alpha\psi\sinh\beta\psi+\frac{\text{E}\alpha+\text{F}\beta}{\alpha^2+\beta^2}\sin\alpha\psi\cosh\beta\psi\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Since $v=0$ when $\psi=0$ the constant of integration in (11) is zero.

The conditions for determining the constants D, E, and F are :

(*a*) When $\psi=\pi$, $v=0$.

(*b*) When $\psi=\pi$, $\dfrac{\partial u}{\partial\psi}=0$.

(*c*) The vertical forces on the ring are in equilibrium, *i. e.*,

$$2\text{RK}\int_0^\pi v\sin\psi\,d\psi=\text{P}_\text{R}.$$

The remaining conditions of equilibrium for the ring are automatically satisfied.

The constants are

$$\text{D}=-\frac{3\text{P}_\text{R}\text{R}^3}{2\text{EI}(1-9\text{X}^2)}\cdot\frac{2\beta}{\sinh 2\beta\pi}\cdot$$

$$\mathrm{E}=\frac{3\mathrm{P_R R^3}}{2\mathrm{EI}(1-9\mathrm{X}^2)(\cosh 2\beta\pi-\cos 2\alpha\pi)}\times\left\{\frac{(6\mathrm{X}-1)\cos\alpha\pi\sinh\beta\pi}{3\beta}+\frac{\sin\alpha\pi\cosh\beta\pi}{\alpha}\right\}.$$

$$\mathrm{F}=\frac{3\mathrm{P_R R^3}}{2\mathrm{EI}(1-9\mathrm{X}^2)(\cosh 2\beta\pi-\cos 2\alpha\pi)}\times\left\{\frac{(6\mathrm{X}-1)\sin\alpha\pi\cosh\beta\pi}{3\beta}-\frac{\cos\alpha\pi\sinh\beta\pi}{\alpha}\right\}.$$

The bending moment at any point in the ring is, from equation (2),

$$\mathrm{M}=\frac{\mathrm{EI}}{\mathrm{R}^2}[\mathrm{D}(\alpha^2+\beta^2)\cosh 2\beta\psi$$
$$-2\beta\{(\beta\mathrm{E}-\alpha\mathrm{F})\cos\alpha\psi\cosh\beta\psi+(\beta\mathrm{F}+\alpha\mathrm{E})\sin\alpha\psi\sinh\beta\psi\}]. \quad (12)$$

The maximum bending moment occurs at the load point, *i. e.* when $\psi=\pi$, and its value is

$$\mathrm{M}_0=\frac{\mathrm{P_R R}}{2(1-3\mathrm{X})}\left[-2\beta\coth 2\beta\pi+\frac{1}{(\cosh 2\beta\pi-\cos 2\alpha\pi)}\times\left\{\frac{2\alpha^2-1}{\alpha}\sin 2\alpha\pi-2\beta\sinh 2\beta\pi\right\}\right]. \quad (13)$$

When Q is very large, $\cosh\beta\pi$ and $\sinh\beta\pi$ are very nearly equal and large compared with $\sin 2\alpha\pi$ and $\cos 2\alpha\pi$. Also $\alpha^2=3\beta^2$ and both α and β are large.

Then $\mathrm{M}_0\rightarrow\frac{\mathrm{P_R R}}{6\beta}$, and in the limit when β is infinite $\mathrm{M}_0=0$. This is the case of an infinitely flexible ring or rigid pins.

When Q is very small, *i. e.* when the ring is very stiff in comparison with the pins, α approaches 1, β approaches zero, and

$$\mathrm{M}\rightarrow\frac{\mathrm{PR}}{2\pi}(1-\tfrac{1}{2}\cos\psi-\psi\sin\psi).$$

The maximum value, at $\psi=\pi$, is

$$\mathrm{M}_0\rightarrow\frac{3\mathrm{PR}}{4\pi}.$$

No useful general expression is obtainable for the position of the maximum value of h, and if this is required it is best found by a process of trial and error.

It occurs where $\frac{\partial v}{\partial\psi}=0$, *i. e.* where $u=0$, and a few values of ψ substituted in equation (10) are sufficient to indicate its position.

In the limiting case, when the ring is rigid, $Q=0$ and

$$h=\frac{P}{\pi R}\sin\psi,$$

which has a maximum value of $\frac{P}{\pi R}$ at $\psi=\frac{\pi}{2}$.

CASE 2.—*Tangential Load at $\psi=\pi$.*

When a tangential load P_T acts at $\psi=\pi$ it is evident from considerations of symmetry that the values of u are of equal magnitude but opposite in sign for equal positive and negative values of ψ.

Hence the constants D, E, and F in equation (9) are all zero, and the solution is

$$u=A\sinh 2\beta\psi+B\cos\alpha\psi\sinh\beta\psi+C\sin\alpha\psi\cosh\beta\psi, \quad . \quad . \quad . \quad . \quad (14)$$

$$v=-\left[\frac{A}{2\beta}\cosh 2\beta\psi+\frac{B\beta-C\alpha}{\alpha^2+\beta^2}\cos\alpha\psi\cosh\beta\psi\right.$$
$$\left.+\frac{B\alpha+C\beta}{\alpha^2+\beta^2}\sin\alpha\psi\sinh\beta\psi+G\right]. \quad (15)$$

The conditions for the determination of the constants are

(*a*) When $\psi=\pi$, $u=0$.

(*b*) When $\psi=\pi$, $M=0$, and since $u=0$, it is clear from equation (2) that $\frac{\partial^2 u}{\partial\psi^2}=0$.

(*c*) The total horizontal force on the ring must be zero, so

$$2RK\int_0^\pi v\cos\psi\,d\psi=-P_T.$$

(*d*) The total moment about the centre of the ring must be zero, so

$$2RK\int_0^\pi v\,d\psi=P^T.$$

These conditions give the following values for the constants :

$$A=-\frac{P_T}{18RK\sinh 2\beta\pi}\frac{(3X-2)(3X+1)}{(3X-1)}.$$

$$B=-\frac{P_T}{18RK}\frac{(3X-2)(3X+1)}{(3X-1)}\left[\frac{(2\alpha^2-1)\sin\alpha\pi\cosh\beta\pi-2\alpha\beta\cos\alpha\pi\sinh\beta\pi}{\alpha\beta(\cosh 2\beta\pi-\cos 2\alpha\pi)}\right].$$

$$C=\frac{P_T}{18RK}\frac{(3X-2)(3X+1)}{(3X-1)}\left[\frac{(2\alpha^2-1)\cos\alpha\pi\sinh\beta\pi+2\alpha\beta\sin\alpha\pi\cosh\beta\pi}{\alpha\beta(\cosh 2\beta\pi-\cos 2\alpha\pi)}\right].$$

$$G=0.$$

The bending moment at any point in the ring is then, from equation (2),

$$M=-\frac{2P_TR}{Q}\frac{\beta^2(\alpha^2+\beta^2)^2}{(4\alpha^2-3)}\left[\frac{\sinh 2\beta\psi}{\sinh 2\beta\pi}-\frac{2}{\alpha(\cosh 2\beta\pi-\cos 2\alpha\pi)}\right.$$
$$\left.\times\begin{Bmatrix}(\alpha\cos\alpha\pi\sinh\beta\pi+3\beta\sin\alpha\pi\cosh\beta\pi)\cos\alpha\psi\sinh\beta\psi\\-(3\beta\cos\alpha\pi\sinh\beta\pi-\alpha\sin\alpha\pi\cosh\beta\pi)\sin\alpha\psi\cosh\beta\psi\end{Bmatrix}\right]. \quad (16)$$

The differentiation of this expression and its equation to zero does not lead to a useful expression for the position of the maximum value of the bending moment. If this value is required it is best to obtain it by a process of trial and error.

The maximum value of h occurs where v is a maximum, *i. e.* where $\frac{\partial v}{\partial\psi}=0.$

Hence the condition for the position of maximum intensity of h is $u=0$, and this is at $\psi=\pi$, *i. e.* at the load point.

Substituting the values of the constants and putting $\psi=\pi$ in the expression $h=Kv$, we obtain

$$h_{\text{max.}}=\frac{P\beta}{R}\frac{(\alpha^2+\beta^2)}{(\alpha^2+3\beta^2)}\left[\coth 2\beta\pi+\frac{3\beta\sin 2\alpha\pi+\alpha\sinh 2\beta\pi}{\alpha(\cosh 2\beta\pi-\cos 2\alpha\pi)}\right]. \quad . \quad (17)$$

When the ring is rigid the general expression for h reduces to

$$h=\frac{P}{2\pi R}(1-2\cos\psi).$$

The maximum value occurs at $\psi=\pi$ and is $\frac{3P}{2\pi R}$. The minimum occurs at $\psi=0$ and is $-\frac{P}{2\pi R}$.

1st August, 1939.

LV. *On Flint's Five-Dimensional Theory of the Electron.*

By William Band, Professor of Physics,
Yenching University, Peking, China *.

[Received May 15, 1939.]

Abstract.

By adopting a different unit for the fifth coordinate from that used by Flint, the fifth component of momentum becomes equal to the spin momentum, $\frac{1}{2}h/2\pi$, and the fifth geodesic equation gives the spin frequency equal to twice the de Broglie frequency. This suggests that the fifth dimension is the angle of spin, the particle spinning through 2π once per half de Broglie wave. The relativity ambiguity of length is now periodic in time with the same period as the de Broglie waves. Flint's deduction of the first-order wave equation is unaffected by the new choice of units.

In a stimulating series of papers, Flint and his co-workers have developed a remarkable application of five-dimensional analysis to the problem of giving a relativistic treatment of the quantum theory.

In five dimensions the track of a charged particle is a null geodesic, analogously with the track of a photon in four dimensions (1). Parallel displacement in five dimensions is free from gauge ambiguity (2), but in four dimensions there appears a variation of length which, however, is periodic in proper time with a period h/mc (3). This is taken as the minimum time observable, so that the ambiguity will not directly affect physical measurements (4). Quantizing the fifth component of momentum can be made to give the charge as a whole number of some unit, taken as the electronic charge (5). A simple gauging equation limiting the changes in matrix length of a five-vector, is shown to be identical with the first-order wave equation of the quantum theory (6).

Perhaps the first point upon which a physicist would feel dissatisfied is that the fifth dimension is still given no direct physical meaning. Flint claims (7) that this difficulty of Kaluza's original theory can be removed by regarding the fifth coordinate as a cyclic coordinate which does not appear explicitly in the final, descriptive, equations of motion.

---

* Communicated by the Author.

But even cyclic coordinates have physical meaning—such as the angle θ of orbital motion quoted by Flint in the last reference; and no such meaning is assigned in Flint's work. In fact, Veblen and Hoffmann (8) take the view that no physical meaning can be given, and therefore prefer to develop the whole theory from a frankly projective point of view which regards the fifth dimension as a mere mathematical convenience. Flint's method certainly has the advantage of simplicity over that of the generalized projective analysis, and hence has a far greater attraction for the physicist who finds Riemannian geometry quite difficult enough!

It occurred to the present writer, who admits being a member of this class of physicist, that another very doubtful point in Flint's work might, on recognition, lead to a much better physical basis for the fifth dimension. The point is that some of Flint's results are not invariant. This is specially serious in his physical interpretation of the fifth component of momentum as the mass of the fundamental particle. In reference (1) above he gives the equation (7):

$$\Pi_5=e/\alpha, \qquad (1)$$

where α is a constant to be chosen at will. (The c is omitted by using $d\tau$ instead of ds in the definition of Π_m.) The only restriction on α is its relation with the other constant γ_{55} through his equation (12) required to make the track a null geodesic

$$\alpha^2\gamma_{55}=e^2/m^2c^2. \qquad (2)$$

By taking $\gamma_{55}=1$, Flint in effect adopts a definite unit in the fifth dimension. We must admit that this is the "natural" unit from a relativistic point of view if we think of x^5 as the same kind of dimension as the others. But it is by no means a natural unit for a cyclic coordinate, as is at once obvious when Flint finds the period of the gauge change to be h/mc in this coordinate.

The natural unit for a cyclic coordinate would be such that the period works out as 2π. To change the unit by the factor $h/2\pi mc$ is easy. We merely have to choose

$$\gamma_{55}=(h/2\pi mc)^2. \qquad (3)$$

and (2) gives

$$\alpha=2\pi e/h. \qquad (4)$$

Flint's expressions for the metric tensor remain unchanged in terms of α; and the geodesic is still null. But the ambiguity in length is now periodic in x^5 with period 2π. The fifth component of momentum, given by (1) above which is still true, instead of equalling the mass of the particle, reduces to

$$\Pi_5=h/2\pi. \qquad (5)$$

This result very forcibly suggests that we should interpret the fifth coordinate as the spin coordinate, and choose the unit so that the ambiguity of gauge has a period 4π in this coordinate, thus giving

$$\Pi_5=\tfrac{1}{2}h/2\pi, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (6)$$

which is the correct angular momentum of the electron spin.

The odd circumstance that the periodicity has to be 4π instead of 2π to produce the spin momentum seems at first unfortunate. But a very surprising result turns up when we consider the angular *velocity* of the spin as given by the fifth geodesic equation. This equation can be written

$$\frac{m}{ds}\gamma_{5\alpha}dx^\alpha=\tfrac{1}{2}h/2\pi c, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (7)$$

see reference (1), equations (11) and (7). This gives

$$m\alpha\gamma_{55}\phi_m dx^m/d\tau+m\gamma_{55}dx^5/d\tau=\tfrac{1}{2}h/2\pi. \quad . \quad . \quad . \quad . \qquad (8)$$

Using our values for the constants in (3) and (4) modified for the period 4π :

$$\gamma_{55}=(h/4\pi mc)^2,\ \alpha=4\pi e/h, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (9)$$

this gives

$$\omega=dx^5/d\tau=\frac{4\pi}{h}\,(mc^2-e\phi_m dx^m/d\tau)$$

or

$$2\pi\nu_{\text{spin}}=4\pi \mathrm{E}/h, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (10)$$

where E is the total energy of the particle.

We therefore find that the proper *frequency* * of the spin is *twice* the frequency of the de Broglie waves representing the particle. The particle, according to this picture, spins once round to every *half* wave, which is, in fact, a more consistent result for a spinning particle which is symmetrical about its axis of spin. There is nothing in the spin angle to associate with the difference between a " crest " and a " trough " ; the spin angle should therefore repeat itself with every node—or half wave.

But the ambiguity of length is now periodic in the spin with period 4π, and hence its period is equal to that of the de Broglie waves. This would seem a much more suggestive relation than the result of Flint's choice of units.

The interesting result that the spin frequency is double the de Broglie frequency is entirely new, arising from the fact that we have in the five-dimensional geodesic equations a method of finding the rotational speed quite independently of any assumptions as to the rotational inertia.

---

* For proper frequencies, $dx^m=(0,\ 0,\ 0,\ \sqrt{-1}\,d\tau)$, so that with $\phi_4=\sqrt{-1}\,\mathrm{V}$, we derive (10) exactly. Since ν and E vary alike in Lorentz transformations, the result holds generally.

Write I for the rotational inertia, then from (10) and

$$\mathrm{I}\omega=\tfrac{1}{2}h/2\pi$$

derive

$$\mathrm{I}=h^2/16\pi^2\mathrm{E}. \qquad (11)$$

Then the rotational energy

$$\tfrac{1}{2}\mathrm{I}\omega^2=\tfrac{1}{2}\mathrm{E}, \qquad (12)$$

or exactly one-half of the total energy of the particle can be ascribed to the energy of its spin.

For a particle at rest in no field $\mathrm{E}=mc^2$, and the radius of gyration r of the particle, defined by

$$\mathrm{I}=mr^2, \qquad (13)$$

will be

$$r=h/4\pi mc. \qquad (14)$$

For the electron this gives $r_{-\mathrm{ve}}=1{\cdot}93\times10^{-11}$ cm. and for the proton $r_{+\mathrm{ve}}=1{\cdot}05\times10^{-14}$ cm. These radii are a hundred times as large as the radii of the spherical electron and proton, which means that the present theory points towards a vortex or disk-shaped particle, if any meaning at all can be assigned to the shape in such cases.

This deduction of I and r is only valid if all the spin energy is mechanical —which means the electromagnetic field is to be neglected. But in this case our values (14) and (10) for r and ω give the linear velocity at the radius r exactly equal to c, the velocity of light. The fact that the "track of an electron" on this theory is a null geodesic thus receives a simple explanation; this is more clearly emphasized when we note that equations (14) and (9) permit us to write

$$\gamma_{55}=r^2. \qquad (15)$$

Thus in the case of no field the five-dimensional expression for square interval becomes

$$d\sigma^2=ds^2+r^2(dx^5)^2. \qquad (16)$$

The four-dimensional interval ds^2 refers to the centre of the particle, and $r\,dx^5$ is simply the further displacement of the point on the disk at radius r. Essentially the expression is not five dimensional at all, in the mysterious sense of adding a fifth unknowable dimension to the four already recognized; the fifth is merely another degree of freedom in the space already known.

Finally, it is worth noting that Flint's deduction of the first-order wave equation is not disturbed by our choice of units. His equation (9)

$$\gamma^5=\alpha_0-\alpha\phi_m\alpha^m$$

should actually read

$$\gamma^5=\alpha_0/\sqrt{\gamma_{55}}-\alpha\phi_m\alpha^m,$$

and with this understanding our units lead at once to the equation deduced by Flint (10), the various changes cancelling out. This was to be expected, since this equation is invariant.

The present interpretation of the fifth dimension is, of course, not consistent with much of Flint's ideas concerning uncertainty relations (4), especially with his deduction of the fundamental unit of electrical charge(5). His quantization rule on Π_5 becomes, on the present interpretation of Π_5, a mere identity with $n=\frac{1}{2}$.

However, it seems well worth while to discard such formal relationships in favour of the present gain in physical meaning.

References.

(1) Proc. Roy. Soc. A, cxxxi. p. 172 (1931).
(2) Proc. Roy. Soc. A, cxxxi. p. 173 (1931).
(3) Proc. Roy. Soc. A, cxxxi. pp. 174, 175 (1931).
(4) Proc. Roy. Soc. A, clix. p. 45 (1937).
(5) Proc. Phys. Soc. l. p. 340 (1938).
(6) Proc. Phys. Soc. l. p. 905 (1938).
(7) Proc. Phys. Soc. l. p. 899 (1938).
(8) Veblen and Hoffmann, Phys. Rev. xxxvi. p. 811 (1930).
Hoffmann, Phys. Rev. xliii. p. 615 (1933).
(9) Reference (6), p. 906, equation preceding (36).
(10) Reference (9), equation (36).

LVI. *The Structure of Meteoritic Iron.*

By Professor E. A. Owen, M.A., Sc.D.,
University College of North Wales, Bangor *.

[Received February 19, 1940.]

[Plate XVI.]

1. *Introduction.*

Some time ago the author published a note (1) in which certain views were put forward concerning the formation of Widmanstätten figures in meteorites. In the present paper it is intended to amplify those views and to modify them where necessary, and in addition to deal generally with the structure of meteoritic iron in the light of work carried out by X-ray analysis in the author's laboratory on a representative collection of meteorites (2) and on artificial iron-nickel alloys rich in iron (3).

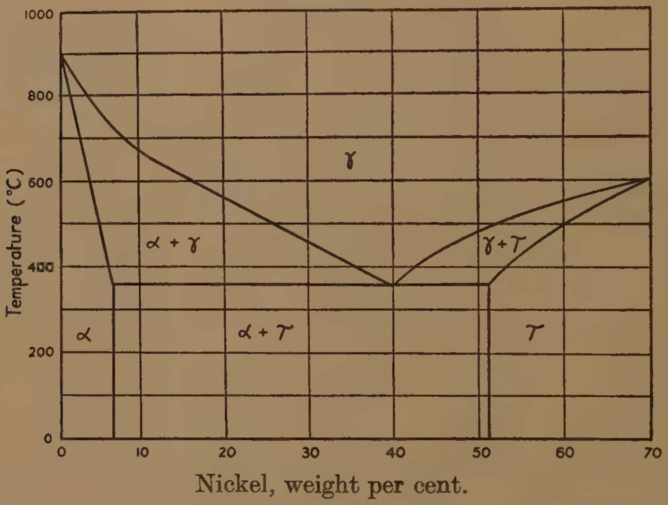
Much work has been done on meteorites and several theories have been advanced to explain their structure. In recent years Mehl and his collaborators have devoted close attention to the Widmanstätten patterns found both in meteorites and in different alloy systems, and they have recorded their results in a series of interesting papers. In the first paper of the series by Mehl and Barrett (4), it is concluded that the first correct explanation of the origin of the Widmanstätten structure was given in 1904 by Osmond and Cartaud (5), according to whom (1) the formation of the Widmanstätten pattern is ascribed to a precipitation of the kamacite (α-phase) along octahedral planes in the original high temperature solid solution (γ-phase), (2) when a solid solution which is stable at high temperatures precipitates a new phase on cooling, this new phase is deposited in such a way that its lattice bears a definite crystallographic relation to the lattice of the parent solid solution. They argued on the basis of the equilibrium diagram of the iron-nickel system shown in fig. 1 which includes a eutectoid transformation at about 350° C. In this diagram also the γ-phase is shown as a separate phase from tænite, the former existing at high temperatures and the latter at temperatures below the magnetic transformation curve, which is treated as a phase boundary. More recent equilibrium diagrams of the iron-nickel system differ in several important respects from this diagram of Osmond and Cartaud; nevertheless, this does not affect the general conclusions which were arrived at by these authors. The second statement has been amply confirmed by various investigators. Young(6), for example, found that the dodecahedral plane (110) of the body-centred cubic kamacite was parallel to the octahedral plane (111) in the face-centred cubic tænite, which was regarded as the residue of the solid solution that was homogeneous at high temperature.

---

* Communicated by the Author.

Other workers who have studied the relative orientations of the α and γ constituents of iron-nickel alloys are Nishiyama(7), Wasserman(8), and Mehl and Derge(9); also Kurdjumov and Sachs(10) have studied the relative orientations of α and γ crystals in a 1·4 per cent. carbon steel.

Fig. 1.



Equilibrium diagram of iron-nickel alloys according to Osmond and Cartaud.

All found a definite relation between the orientations of these two components. With iron-nickel alloys, including meteoritic iron-nickel, the results of Young were confirmed, but other relative orientations in addition to that found by Young are mentioned; for instance, Nishyama finds also [211] γ parallel to [110] α, and this was confirmed by Wassermann; Kurdjumov and Sachs find in carbon steel that in addition to the (111) γ being parallel to the (110) α; [110] γ is also parallel to [111] α, that is, the [110] direction in any particular (111) γ-plane is parallel to a [111] direction in the corresponding (110) α-plane.

Mehl and Derge(9) were able to produce in an artificial iron-nickel alloy containing 27 per cent. nickel, patterns similar to the Widmanstätten figures found in meteorites. The alloy was annealed in hydrogen at about 1400° C. for 2 weeks to allow the growth of large grains of the γ-phase; it was then cooled in the furnace to room-temperature in 12 hours. The relative orientations of the α- and γ-components in this sample were found to agree with the relation of Kurdjumov and Sachs. Another alloy containing 31 per cent. nickel and given the same initial annealing treatment, cooled to room-temperature and then annealed at 240° C. for 6 days yielded a similar pattern, in which the relative orientations of the α- and γ-constituents were the same as in the previous specimen.

From a study of the meteorites Cañon Diablo and Amalia Farm, Derge and Kommel (11) conclude that the structures found in metallic meteorites were formed during slow cooling from high temperatures. They refer to the artificial iron-nickel alloy containing 27 per cent. nickel mentioned above, and state that, although this alloy contained more nickel than the average meteorite, there is little reason to doubt that this example provides an illustration of the production of the Widmanstätten pattern which was first observed in meteorites. They are of the opinion that it is necessary to maintain the temperature at about 1400° C. in order to grow large γ-crystals which, it is assumed, are required for the formation of the Widmanstätten pattern. Their picture of the changes taking place is contained in the following paragraph quoted from their paper: "Below approximately 1400° C., the entire mass will be in the face-centred cubic γ form and these alloys have a pronounced tendency to form large grains at these temperatures, thus accounting for the huge tænite crystals required for the formation of the Widmanstätten pattern. On further cooling, as the alloy reaches the γ-($\alpha+\gamma$) boundary, it will begin to transform to the body-centred cubic α-phase, forming the kamacite plates. During this process the remaining γ will become richer in nickel until it contains about 25 per cent.; material containing more nickel than this does not transform readily at ordinary temperatures, and will be the component known as tænite. Some of this tænite will have a composition in the two-phase field ($\alpha+\gamma$) and will therefore transform slowly on further cooling, either partially or completely, depending upon its composition and rate of cooling, giving rise to the various forms of plessite. It is believed that all structures found in metallic meteorites can be explained on this general theory."

This statement represents the view at present generally accepted concerning the formation of the Widmanstätten structure, but there are several points in the statement concerning which more definite information would be desirable. It should be mentioned also that the authors base their arguments on the iron-nickel equilibrium diagram put forward by Marsh (12) which, according to more recent work, requires modification.

Smith and Young have devoted careful attention to the Widmanstätten structures found in octahedral meteoritic iron and have recently published a summary of their results (13). These will be referred to later in this paper. They conclude from their researches that it is highly probable the kamacite lamellæ are the result of a $\gamma \rightarrow \alpha$ transformation.

The theory of the structure of meteoritic iron which will be explained in this paper differs materially from that referred to above. Even if it may not supply the complete solution it is worthy of consideration, as it provides an explanation of certain aspects of the problem and of the experimental results obtained with some artificial iron-nickel alloys which existing theory fails to give.

Before proceeding, however, it would be well to consider the more recent equilibrium diagrams suggested for the iron-nickel system and some of the recent results obtained with quenched artificial alloys and with meteorites.

2. *Equilibrium Diagram of Iron-Nickel Alloys.*

The equilibrium diagram for iron-nickel alloys, rich in iron, put forward by Bradley (14) is different in many respects from the equilibrium diagrams already advanced for this part of the system. He uses the diagram to explain certain aspects of the structure of meteorites, and concludes that the typical meteoritic structure is only in equilibrium between 350° C. and 580° C., and not below 350° C. Such meteorites must therefore have received their heat treatment above 350° C. They must also have cooled too quickly for equilibrium to be followed down to terrestrial temperatures, and at these temperatures they are too sluggish for further changes to take place.

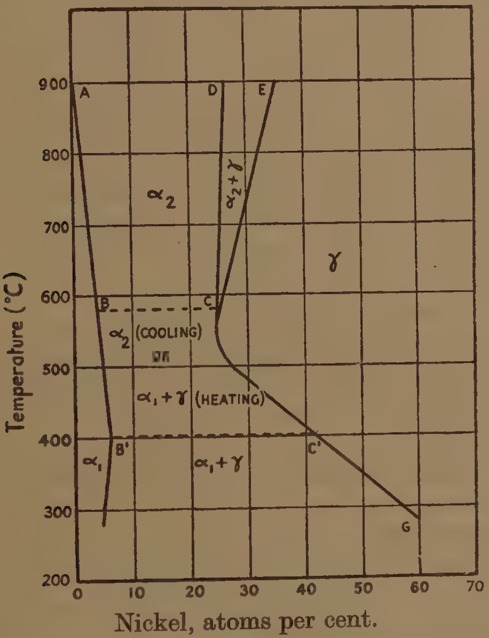
Bradley considers that the most effective part of the heat treatment a meteorite receives by its proximity to the sun is that which takes place in the $(\alpha+\gamma)$ range, that is, between the tempratures of 580° C. and 350° C. As the crystals grow it becomes increasingly difficult for the 350° C. transformation to take place.

Owen and Sully (3) show two diagrams in their paper on artificial iron-nickel alloys, one representing equilibrium conditions and the other, reproduced in fig. 2, representing the results obtained with quenched alloys, some of which were not in their final state of equilibrium.

For details of the heat treatments given to the ingots and to the powder specimens reference should be made to the original paper. Above about 580° C. it did not matter whether the final temperature was reached from above or from below this temperature ; the alloys on quenching always yielded a body-centred cubic structure, which was distorted and contained excess of nickel. The material may be regarded as a supersaturated solid solution of nickel in iron, which breaks up into a mixture of α- and γ-phases in equilibrium when annealed again at temperatures below 580° C.

Over a range of temperature between about 580° C. and 400° C.* the quenched alloys showed different structures according as the temperature was lowered or raised to the final temperature. If lowered, the distorted α_2 form appeared; if raised, both α- and γ-phases were present together. This is an instance of the irreversibility which is usually associated with iron-nickel alloys extending over certain ranges of composition and temperature.

Fig. 2.



Phase diagram of iron-nickel alloys showing spurious boundaries obtained with quenched specimens.

3. *Structure of certain Meteorites.*

A full account of the X-ray examination of 24 iron meteorites in the natural state and after being annealed at different temperatures, together with the parameters of the constituent phases (kamacite and tænite) is included in a previous paper (2). The results are summarized below.

---

* This temperature was given as 430° C. in a previous paper (3), but more recent work has brought it down to 400° C., and it may still be lowered. The difficulty of fixing the temperature definitely is due to the length of time required for annealing at these low temperatures.

(i.) The meteorites, which included hexahedrites, octahedrites, mesosiderites and ataxites, ranged in nickel content from about 4·4 to 25·6 per cent. After annealing for periods of 5 or 8 weeks at 350° C., they consisted of either the α-structure, if the nickel content was less than 6·2 per cent., or a mixture of α- and γ-structures if the nickel content was higher than this.

(ii.) The spectra of the α-phase in the annealed specimens contained well-defined lines from which accurate values of the parameters could be obtained. The α-parameter was found to increase linearly with increasing nickel content in the pure α-phase, and to remain approximately constant in the duplex region, the mean value being $2{\cdot}8628_2$ Å. The greatest deviation from this value was 1 in 20,000. For the 12 octahedrites contained in the group, the greatest deviation from the mean value was 1 in 30,000. From the constancy of the α-phase parameter values, it was concluded that the mean composition of kamacite in octahedrites is the same whatever the width of the kamacite band found in the Widmanstätten pattern, also that the composition of the kamacite component of plessite is, within the limits of measurement, the same as that of the kamacite bands.

(iii.) The definition of the γ-lines was in general not so good as that of the α-lines, and in some meteorites was very poor. The variation in the γ-parameter value from one meteorite to another was quite marked and seemed to bear no relation to the nickel content of the meteorite. The relative variations in the α- and γ-parameters for the ataxites examined agreed well with the relative variations to be expected on the basis of the equilibrium diagram of the artificial alloys (3).

(iv.) When annealed at different temperatures ranging from 350° C. to 500° C., (*a*) the hexahedrites, containing the α-constituent alone, showed no variation beyond the limits of error of measurement. This is in accordance with expectation, since these meteorites have compositions falling within the range of the pure α-phase which is the region of unsaturated solid solution of nickel in iron. (*b*) The octahedrites ranging in nickel content from 7·26 to 13·18 showed remarkably constant α-parameters at all temperatures. The γ-lines were too faint to measure with accuracy in most of the octahedrites. (*c*) The ataxites, rich in nickel, showed the same variation of α-parameter as that observed in artificial alloys of about the same composition and similarly heat treated.

(v.) Unannealed material from the hexahedrites and octahedrites yielded spectra identical with those obtained after annealing at temperatures between 350° C. and 500° C. The heat treatments had not appreciably changed the state of the meteorites. With ataxites that contained between 17·3 and 25·6 per cent. nickel the structures in the unannealed state were different from those present in the annealed state. Whereas the annealed specimen contained both the α- and the γ-phases, the unannealed material contained only α in a distorted form. This was identified with the metastable α_2-structure (martensite) found in artificial alloys when quenched from high temperature. Three of these ataxites were examined in detail, namely, Babb's Mill, Klondyke, and San Cristobal containing respectively 17·30, 18·2, and 25·6 per cent. nickel. In the unannealed state these meteorites, which should be richest in tænite, did not show any γ-lines in their spectra ; but it was found that an annealing treatment of only 8 days at 300° C. was sufficient in each case to change the α_2-structure into the $(\alpha+\gamma)$ state, that is, to transform martensite into plessite. An annealing treatment of 56 days at 250° C. brought about a similar change in two of the meteorites. These results point to the conclusion that the equilibirium condition at these low temperatures over the range of composition from 17 to 25·6 per cent. nickel is that in which the α- and γ-phases coexist. The cold work done in producing the filings seems to have a marked effect on the structure of the material. It should be emphasized that further investigation is necessary before it is safe to draw final conclusions from the limited number of experiments carried out with unannealed specimens.

(vi.) Generally it was found that meteoritic irons closely resemble artificially prepared iron-nickel alloys in their reaction to different heat treatments. It was concluded that the three meteorites, Babb's Mill, Klondyke, and San Cristobal behaved when subjected to heat treatment as would be expected if they had been rapidly cooled from elevated temperatures.

4. *The Existing Theory of the Structure of Meteoritic Iron.*

The conclusions reached by previous workers concerning the structure of meteoritic iron may be summarized as follows ; they form the basis of the existing theory : (*a*) Large γ-crystals are required for the formation of the Widmanstätten pattern. (*b*) The structures found in meteorites are produced during slow cooling from high temperatures at which large γ-crystals are formed. (*c*) The kamacite lamellæ are the result of a $\gamma \rightarrow \alpha$ transformation in accordance with the original suggestion of Osmond and Cartaud.

It is not clear how the different forms of plessite are produced. The explanation given by Derge and Kommel does not appear to be satisfactory.

Figures resembling the Widmanstätten figures found in meteorites have been produced in artificial iron-nickel alloys. These are formed either by cooling the alloy to room-temperature from high temperature at a comparatively rapid rate, namely, from 1400° C. to room-temperature in 12 hours, or by annealing the material after cooling, *e. g.*, at 240° C. for 6 days. From these experiments it would appear that exceedingly slow cooling is not necessary.

5. *An Alternative Theory of the Structure of Meteoritic Iron.*

(i.) *Basic Assumptions.*

An attempt will be made here to explain the formation of the structures found in meteorites on a theory based on the view that the transformations operating are $\gamma \rightarrow \alpha_2 \rightarrow (\alpha_1 + \gamma')$, that is, the γ-phase existing at high temperatures is transformed into the metastable α_2-form, which on annealing at comparatively low temperatures transforms into a mixture of α- and γ-phases in equilibrium. The gamma-phase is designated in the above transformation as γ and γ' to indicate different compositions, although the structure is the same. The theory does not preclude the possibility that the material may initially be in the molten state and by rapid cooling pass firstly into the γ-phase and then into the α_2-state. It may also include rates of cooling which allow the transformation $\gamma \rightarrow \alpha_2$ to take place with enough time on further cooling for the material to change over from α_2 to $(\alpha_1 + \gamma')$.

It will be noted that, in contrast to the assumption that large γ-crystals exist in the meteorite initially, the suggestion now is that whatever the size of the γ-crystal originally, it is transformed into a mass of small crystals in the α_2-state, the orientations of which may vary amongst themselves within narrow limits and on the average may bear a relation to the orientation of the parent γ-crystal. If the original γ-crystals are small, the orientations of the α_2-crystals will appear to be random, but if the parent γ-crystal is large, the α_2-crystals build up into an imperfect crystal the orientation of which will be related to that of the original γ-crystal. When annealed at moderate temperatures these α_2-crystals transform into the two stable components kamacite and tænite, portions of which grow in certain directions to form large crystals, and these will enclose regions in which crystal growth has not been so pronounced, resulting in the formation of plessite.

(ii.) *How the Meteorite may get into the α_2 State.*

In the note already referred to (1), it was suggested that for the formation of the Widmanstätten pattern, the meteorite must be brought at some time in its existence to the α_2-condition. The suggestion that this takes place when the meteorite reaches the earth needs to be modified in view of the fact that the internal temperature of the meteorite may not be high enough to produce the necessary state on quenching. Furthermore, even though the temperature of the interior were high enough, and there were lattice distortion due to the excess of nickel present in the α_2-phase, and also mechanical distortion due to the impact with and penetration into the earth's crust, there is not enough time for the crystals to form to the perfection found in octahedrites. Another fact which militates against the view that the figures are formed after reaching the earth is the distortion found in some meteorites, notably the Henbury irons. Probably this distortion took place during the impact of the meteorite with the earth; the Widmanstätten pattern was present before the disturbance occurred. Smith and Young (15) are also of the opinion that the view expressed in the note is untenable.

The point, however, does not affect the theory now advanced in which it is assumed that the initial γ-state of the material is changed directly into the α_2-state by rapid cooling. Rather than that the α_2-state should be formed when the meteorite reaches the earth, it may be assumed to be formed in the early stages of the existence of the meteorite.

A recent review of the few available observations of meteorite trajectories indicates that the observations are not sufficiently reliable to justify the conclusion which a number of writers have reached, that the majority of meteorites originated outside the solar system. At present there does not seem to be sufficient data available to justify any definite decision as to the origin of meteorites. Farrington (16), however, favours the view that meteorites are the result of the disintegration of bodies within the solar system, and some recent observations on meteorite trajectories also indicate that most of the large masses which reach the earth's surface as meteorites originate within the solar system, having elliptic orbits. Chamberlin (17) suggests that a probable cause of disruption may be found in the differential attraction produced by the passage of a small body within a certain distance to a larger dense one. The passage of a large body near a small one would also tend to change the orbit of the smaller body.

It is assumed that after disruption, the hot body, whether molten or solid, is suddenly cooled, resulting in the production of a meteorite in the α_2-condition. It will consist of a supersaturated solution of nickel in iron, in which there will be internal stresses caused by the distortion of the crystal lattice owing to the excess of nickel, and also by the distortion due to mechanical causes. Such material is in an unstable condition, and if heated to moderate temperatures will break-up into the stable components kamacite and tænite found in meteorites (3).

(iii.) *Metakamacite.*

The α_2-material will be the granular structureless form of plessite as distinguished from the form of plessite which contains fine crystals of kamacite and tænite in octahedral arrangement. On the view taken here these two forms of plessite are fundamentally different, and should be called by different names. For the purpose of this paper the form consisting of fine crystals of kamacite and tænite will be referred to as plessite, and the form consisting of the granular α_2-material will be called metakamacite *.

(iv.) *The Plessite Areas in Widmanstätten Figures.*

After the initial formation of metakamacite, the meteorite undergoes heat treatment at temperatures which, according to the results obtained with artificial iron-nickel alloys (3), should not exceed about 580° C. If the meteorite settles down to a definite orbit, it will be alternately heated and cooled gradually as it moves towards or recedes from the sun. In some meteorites, as for instance the octahedrites, this may take place a large number of times (14). Since metakamacite is metastable and is subjected to internal stresses, it is in a condition which is favourable to the growth of large crystals when the material is annealed. As the annealing progresses the crystals grow rapidly in certain directions, and the internal stresses disappear before the crystals in certain parts of the material have had time to show much development; in a plane section of the material these will appear as the plessite areas. On this view it is to be expected that the areas enclosed by the large crystals should consist of a mixture of small crystals of kamacite and tænite, that is, plessite and not metakamacite. In the octahedrites examined in detail during the course of the investigation, this appeared to be the case, but more specimens should be examined before coming to a final conclusion on this point.

(v.) *Process of Crystal Growth.*

During the process of crystal growth, the surplus nickel in the metakamacite lattice is gradually displaced out of solution leaving the pure α-lattice (kamacite), which grows in sheets parallel to the dodecahedral planes, the growth at right angles to this direction being small †. The nickel collects at the surfaces of the kamacite sheets, and, when the concentration at the boundary reaches a certain value, tænite in equilibrium with the kamacite is formed. Thus tænite appears in the form of very thin sheets bordering crystals of kamacite and separating them from plessite areas. By this mode of formation the tænite band will be very narrow, and its nickel content will vary from the equilibrium value to lower concentrations of nickel in passing across it from the kamacite band to the plessite area. This, in addition to the fact that the temperature of annealing may not be constant, and that the equilibrium composition of tænite varies more rapidly with temperature than that of kamacite, will account for the diffuse lines always observed in spectra of tænite.

---

* The α_2-structure found in iron-nickel alloys was previously referred to as martensite (Phil. Mag. xxvii. p. 618 (1939)), following the terminology employed in carbon steels, but it is desirable to give it a different name, as its crystal structure is cubic whereas that of martensite is tetragonal.

† *Cf.* J. Young, *ibid.* p. 640.

Thus the tænite layer on the side bordering the kamacite will be well defined, whilst on the side bordering the plessite area it will present an irregular outline. Two such tænite layers formed close to each other will result in what appears to be a wide tænite band less rich in nickel at its centre than at its boundaries. Such parallel layers of tænite can be seen in the Widmanstätten figures shown in figs. 3 and 4 (Pl. XVI.) *. Where the layers are not very close to one another, the enclosed area consists of plessite formed out of metakamacite, the region being one where there has been no abnormal crystal growth. Smith and Young (13) refer to the plessite enclosed within a broad tænite band as being due to an independent $\gamma \rightarrow \alpha$ transition.

(vi.) *Relative Orientation of Kamacite and Tænite Crystals.*

It may be expected as the result of the process of crystal growth outlined above, that a definite relation should exist between the orientation of kamacite and tænite crystals such as that which has been found by different observers, and to which reference has already been made. The relative orientations of the two components should also be the same in the plessite areas as in the band structures.

(vii.) *Different Width of Kamacite Bands.*

The lower the nickel content of the octahedrite the greater is the thickness of the kamacite sheet, because a greater volume of kamacite will have to be formed in order to expel a sufficient number of nickel atoms to produce tænite of the required equilibrium composition at the boundary. This explains the different degrees of " coarseness " found in octahedrites.

(viii.) *Mosaic Pattern in Band Structures.*

Reference has already been made to figs. 3 and 4. Fig. 3 shows the Widmanstätten pattern found in the meteorite Boogaldi containing 8·05 per cent. nickel. In this photograph, which is of low magnification (× 5), the wider dark bands are kamacite, and the fine white lines bordering them and the plessite areas are tænite. Sufficient detail is contained in the photograph to show up in places the Neumann lines and also the irregular breaking up of the kamacite bands. These features are shown perhaps more clearly in fig. 4, which is a photograph taken at higher magnification (×11) of a polished and etched surface of the meteorite Mantos Blancos, containing 8·83 per cent. nickel. In this photograph the fine white bands are tænite; they border the dark plessite areas and the much wider but lighter coloured bands of kamacite. Here the tænite bands appear continuous except for a few fine markings, whereas the kamacite bands are broken up into small crystals exhibiting a mosaic pattern; these somewhat irregular crystal boundaries are in addition to the more regular markings due to Neumann lines.

---

* I am indebted to Mr. B. D. Burns, M.Sc., of this laboratory, for taking these photographs.

(ix.) *Effect of the Unequal Thermal Expansions of Kamacite and Tænite.*

It has been found that meteoritic irons closely resemble artificially prepared iron-nickel alloys in their behaviour under different heat treatment (3). Also the measurement of the thermal coefficients of expansion of the α- and γ-components of iron-nickel alloys, rich in iron, showed that the γ-component expanded more rapidly than the α-component (18).

On account of the variation of temperature experienced by the meteorite in its motion round the sun and the disparity in the thermal expansions of the α- and γ-components, stresses will be set up in the material. Each band of kamacite is bordered as we have seen by layers of tænite which are thin compared with the thickness of the kamacite band. The effect of the alternate heating and cooling to which the meteorite is subjected and the unequal expansions of the two constituents kamacite and tænite, will be, firstly, to aid the growth of the crystals by setting up internal stresses and, secondly, to tend to break-up the crystals after they are formed and to produce slip bands; this effect will be more evident in the wide kamacite bands than in the thin tænite layers. The mosaic pattern and the fine lines found in the laminar structures of octahedrites, especially in the wider kamacite bands, may be explained in this way. If this view is correct it would not be surprising to find that kamacite crystals in the band structures exhibit a certain variation of orientation, whereas the orientations of the tænite crystals are more uniform in relation to the kamacite-tænite boundary.

The theory as above outlined explains the main features of the Widmanstätten patterns found in octahedrites.

6. *Ataxites.*

The ataxites are considered to be bodies which have not been submitted to prolonged heat treatment after the initial cooling or, if they have, their structure was partly, if not entirely, destroyed by excessive heating before again being rapidly cooled. When they came within the sphere of the earth's attraction either they were already in the metakamacitic state or they were at high temperature and were rapidly cooled, so that they reached the surface of the earth mainly in the metakamacitic condition. Instances of ataxites, which are mainly in the metakamacitic state, are Babb's Mill, Klondyke, and San Cristobal (2). It is desirable to examine many more of these nickel-rich ataxites to test the above assumptions and to arrive at an explanation, if possible, why no meteorite with more than about 15 or 16 per cent. of nickel has fully developed duplex structure, and to investigate if there exist different degrees of metastability according to the nickel content, with a view to throwing light on the difference in the development of the duplex structure generally in meoritic irons.

7. *Some Previous Results in the Light of the above Theory.*

During the X-ray examination of a number of octahedrites it was found, as already stated, that the parameter of tænite showed much greater variation than that of kamacite. This disparity between the variation of the parameters of the two components is in accordance with the equilibrium diagram proposed for the iron-nickel system from a study of artificial alloys if it is assumed that the meteorites have undergone annealing treatments at different mean temperatures. The values of the γ-parameters were such that the mean temperatures of annealing of the particular meteorites examined fell between about 550° C. and 300° C.

Smith and Young (13) have investigated the Butler meteorite which contains about 10 per cent. nickel. By measuring the magnetic transformation temperature of tænite *in situ* in the meteorite, it was found to be about 500° C., which indicates that the material must contain approximately 50 per cent. of nickel, and, according to the equilibrium diagram of Owen and Sully (3), this corresponds to the boundary composition of tænite at about 350° C. The meteorite had therefore been maintained at a mean temperature of 350° C. for a period long enough to allow it to reach a state of equilibrium at this temperature. The laminar structures and plessite areas found in the meteorite would be formed on the present theory from metakamacite in the manner already described.

On the $\gamma\rightarrow\alpha$-transformation theory the meteorite is supposed to have been held initially at high temperature (about 1400° C.) to allow the growth of large γ-crystals; it is then cooled so slowly that equilibrium conditions between kamacite and tænite are maintained down to 350° C., when the rate of cooling becomes too fast for equilibrium conditions to obtain. It is difficult to see how starting with large crystals plessite areas are formed during the process of slow cooling. If the plessite is formed during the final cooling from 350° C. then its composition would be that of tænite at the boundary and would be richer in nickel than the original tænite. But this does not appear to be the case.

Mehl and Barrett (4), when referring to the work of Young on the relative orientation of kamacite and tænite crystals in meteorites, state that the tænite may be regarded as being the undecomposed residue of the high temperature phase. This, of course, cannot mean that the tænite bands found in meteorites have the same composition as the original γ-crystals formed at high temperature, because if the transformation γ to α is supposed to take place so slowly that equilibrium conditions obtain throughout the process, the composition of the γ-crystals changes to correspond with that at the boundary until the change in temperature becomes too rapid for the composition further to adjust itself to it. In the Butler meteorite referred to above the composition of the tænite found in it differed very greatly from that of the original tænite existing at high temperature.

Smith and Young (13) found that small random deviations from the orientation relationships, namely (110) kamacite parallel to (111) tænite, and [001] kamacite inclined at 4° to $[01\bar{1}]$ tænite, are found in the kamacite crystals, but that the majority of the tænite crystals have a single orientation throughout the meteorite, their {111} planes lying nearly parallel to the kamacite-tænite lamellæ. They regard this as supporting the view that the kamacite lamellæ are the result of the $\gamma \rightarrow \alpha$ lattice mechanical transformation, and this is considered as being confirmed by the fact that approximately the same relationship is found in a 10 per cent. artificial alloy in which the $\gamma \rightarrow \alpha$ transformation was known to have occurred.

On the present theory the variation in the orientations of the kamacite crystals and the single orientation found in the tænite crystals is explained by the breaking up of the kamacite bands into small crystals, if they are not already present, on account of the differential expansion and contraction of the two constituents when the meteorite is heated and cooled, combined with the fact that the tænite sheets are very thin compared with the kamacite bands.

As to the artificial alloy containing 10 per cent. nickel, Smith and Young state that it was cooled slowly from the molten state and that when examined by magnetic and X-ray analysis it was found to consist entirely of body-centred crystals (13). It is clear from this that the cooling had not been sufficiently slow to maintain equilibrium conditions, and that the material had been transformed into the metastable α_2-state. They found that when the alloy was heated progressively to and annealed at a temperature of 580° C., its subsequent behaviour indicated the presence of a γ-phase alloy containing about 30 per cent. nickel *. The authors state that X-ray examination confirmed the presence of γ-crystals showing "a strongly preferred orientation, in which their {111} planes were parallel to the Widmanstätten figures in each grain, as in the meteorites. The α-crystals also exhibited twenty-four orientations differing little from those found in the octahedral meteorites."

As it appears that the material was in the metakamacitic state before it was finally annealed, the above observations lend support to the theory outlined in this paper rather than to the existing theory.

I wish to thank Dr. Max H. Hey, of the Mineralogical Department, British Museum, for much useful information concerning meteorites and for valuable criticisms and suggestions concerning the subject-matter of the paper.

References.

(1) E. A. Owen, 'Nature,' cxlii. p. 999 (1938).
(2) E. A. Owen and B. D. Burns, Phil. Mag. xxviii. p. 497 (1939).
(3) E. A. Owen and A. H. Sully, Phil. Mag. xxvii. p. 614 (1939).
(4) R. F. Mehl and C. S. Barrett, Trans. Am. Inst. Min. Met. Eng., Iron and Steel Division Preprint, Tech. Publ. no. 797 (1937).
(5) F. Osmond and G. Cartaud, *Rev. de Mét.* i. p. 69 (1904).
(6) J. Young, Proc. Roy. Soc. cxii. p. 630 (1926).
(7) Z. Nishiyama, Sci. Rep. Tôhoku Imp. Univ. ser. i. xxiii. p. 637 (1934).
(8) G. Wassermann, *Mitt. Kaiser-Wilhelm Inst. f. Metallforsch*, xvii. p. 149 (1935).
(9) R. F. Mehl and G. Derge, Trans. Am. Inst. Min. Met. Eng., Iron and Steel Division Preprint, Tech. Pub. no. 797 (1937).
(10) G. Kurdjumow and G. Sachs, *Zeits. f. Phys.* cxiv. p. 325 (1930).
(11) G. Derge and A. R. Kommell, Am. Journ. Sci. xxxiv. p. 203 (1937).
(12) J. S. Marsh, 'Alloys of Iron and Nickel,' i. p. 54 (1938 ed.).
(13) S. W. J. Smith and J. Young, 'Nature,' cxliii. p. 384 (1939).
(14) A. J. Bradley, 'Nature,' cxliii. p. 518 (1939).
(15) S. W. J. Smith and J. Young, 'Nature,' cxlii. p. 1162 (1938).
(16) O. C. Farrington, 'Meteorites,' p. 211 (1915 ed.).
(17) T. C. Chamberlin, Astrophysics Journn. xiv. p. 17 (1900).
(18) E. A. Owen and E. L. Yates, Proc. Phys. Soc. xlix. p. 178 (1937).

---

* An artificial alloy containing 10 per cent. nickel submitted to this heat treatment would, according to the results of Owen and Sully, contain a mixture of kamacite (containing about 4 per cent. of nickel) and tænite (containing between 15 and 20 per cent. nickel).

LVII. *Note on a Method for the Comprehensive Analysis of Complex Power Plant.*

By A. C. Walshaw, M.Sc., D.I.C. (Lecturer in Applied Thermodynamics at the Northampton Polytechnic, London) *.

[Received January 19, 1940.]

This note is the result of further investigation into the ideas and method presented in a previous paper (Phil. Mag. ser. 7, vol. xxvi. p. 784 (Nov. 1938)), and shows how an analysis may be made of the various quantities of energy flowing along the interconnected streams of any power plant. The analysis results in a series of equations connecting rates of flow of energy along the several power streams, by means of which balance sheets and energy flow diagrams can be drawn up for any part of or the whole of any power plant.

The particular case chosen for analysis here is that of a six-cylinder, supercharged, compression-ignition oil-engine plant—the supercharger being driven by an exhaust gas turbine. The accompanying diagram represents the plant, and it will be seen that the whole has been enclosed by an outer boundary within which there are several inner boundaries enclosing individual units which together make up the whole plant. The individual units are all either energy transformers dealing with heat, electrical energy, kinetic and mechanical energy, or units in which transference of heat takes place. The quantities of energy flowing per unit of time along the various streams entering and leaving the several boundaries have been represented by letters with an attendant arrow to indicate direction. The following energy or power equations for the individual units and also for the whole plant can be deduced from the fact that, when the plant is running under steady conditions, the energy per unit of time entering any particular boundary must be equal to the energy per unit of time leaving that same boundary.

Engine unit. $F_0+L_i+J_i+A_0=M+L_0+X+J_0+B+r_e$.

Fuel pump. $M+F_i=F_0+r_p$.

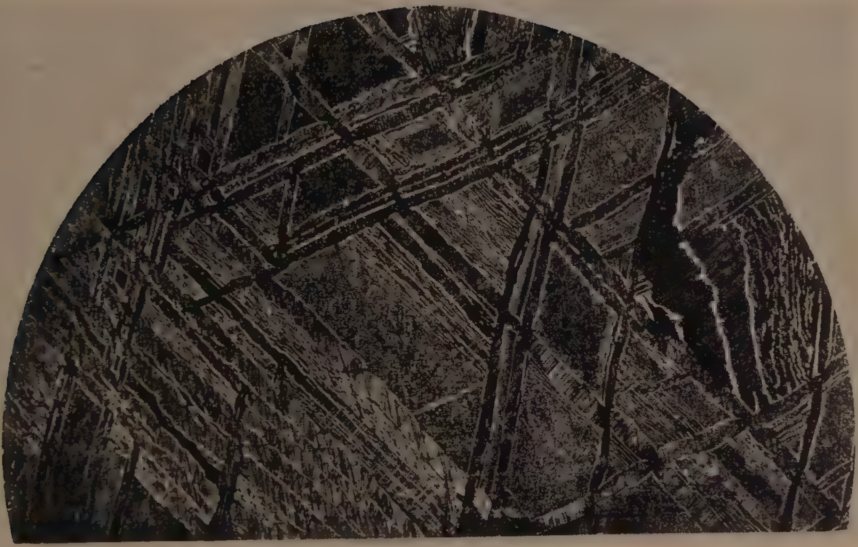
Lubricating-oil pump. $e_0+L=L_i+r_L$.

Lubricating-oil cooler. $W_i+L_0=L+W_0+r_c$.

---

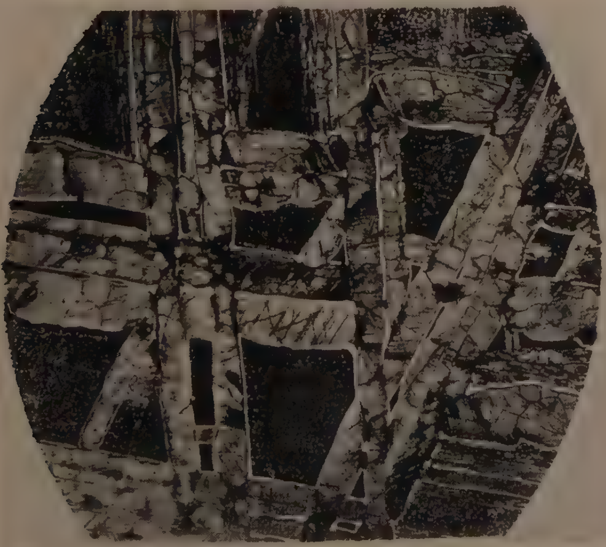
* Communicated by the Author.

Fig. 3.



Widmanstätten pattern found in the meteorite Boogaldi. (×5.)

Fig. 4.



Widmanstätten pattern found in the meteorite Mantos Blancos. (×11.)

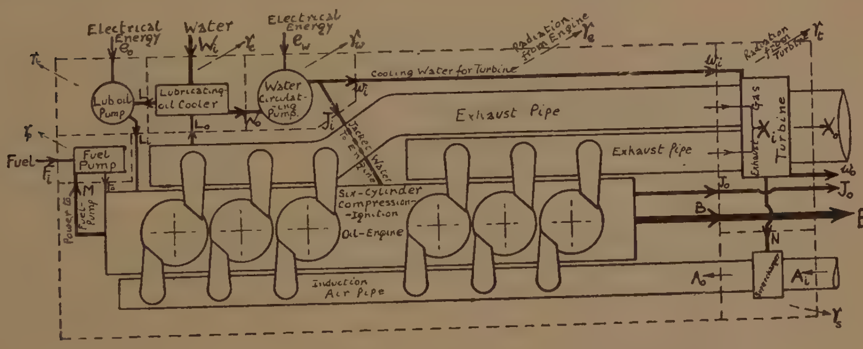
Water pump.	$e_w+W_0=J_i+w_i+r_w.$
Exhaust gas turbine.	$X_i+w_i=N+X_0+w_0+r_t.$
Supercharger.	$N+A_i=A_0+r_s.$

These equations can be rearranged so as to facilitate the elimination of certain terms when the seven equations are added together, as follows :—

$$\begin{array}{lcl}
0 &=& B-F_0+M-L_i+L_0-J_i+J_0+X_i-A_0+r_e,\\
F_i &=& +F_0-M \quad +r_p,\\
e_0 &=& -L+L_i \quad +r_L,\\
W_i &=& -L+W_0-L_0 \quad +r_c,\\
e_w &=& -W_0+w_i+J_i \quad +r_w,\\
0 &=& X_0 \quad -w_i+N+w_0-X_i \quad +r_t,\\
A_i &=& -N \quad +A_0+r_s.
\end{array}$$

Thus, $$F_i+e_0+W_i+e_w+A_i=B+X_0+(J_0+w_0)+\Sigma r. \quad . \quad . \quad . \quad \text{(I.)}$$

This last equation, as will be seen on referring to the diagram, is the energy equation which can be deduced for the whole plant enclosed within


Electrical Energy
e_0
Water
W_i
r_c
Electrical Energy
e_w
r_w
Radiation from Engine
r_e
Radiation from Turbine
r_t
r_L
r_p
Fuel
F_i
Lub oil Pump
L
Lubricating oil Cooler
L_0
L_i
W_0
Water Circulating Pump.
w_i
J_i
Cooling Water for Turbine
w_i
Exhaust Pipe
Exhaust Pipe
Jacket Water to Engine
Fuel Pump
F_0
M
Power to Fuel-Pump
Six-Cylinder Compression-Ignition Oil-Engine
Induction Air Pipe
Exhaust Gas Turbine
X_i
X_0
w_0
J_0
J_0
B
B
N
Supercharger
A_0
A_i
r_s

the outer boundary ; and the fact that it has been obtained by the merging of the equations deduced from the individual units composing the plant provides a check on the accuracy of all previous equations.

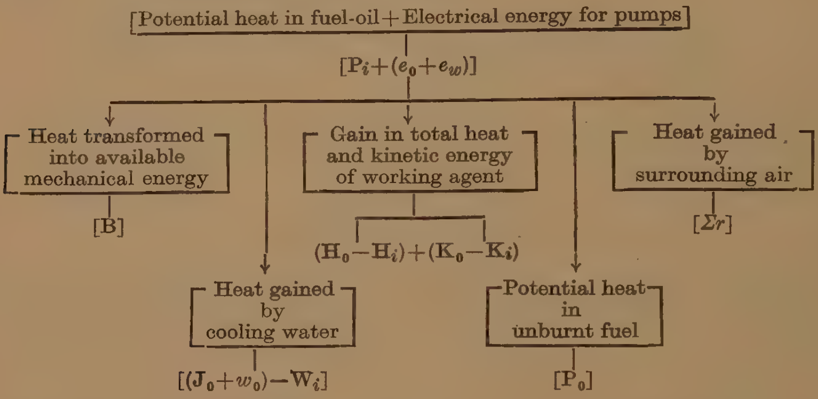
Since the fuel (F_i) and air (A_i) entering the boundary and also the exhaust gases (X_0) leaving the boundary have potential energy (P), kinetic energy (K), and total heat (or enthalpy, H), the difference between the energy of the working agent when entering and leaving the outer boundary is represented by the equation

$$[(F_i+A_i)-X_0]=(P_i-P_0)+(H_i-H_0)+(K_i-K_0). \quad . \quad . \quad . \quad \text{(II.)}$$

Thus, on substituting equation (II.) in (I.), we get the resultant energy equation :

$$[P_i+(e_0+e_w)]=B+[(J_0+w_0)-W_i]+(H_0-H_i)+(K_0-K_i)+P_0+\Sigma r \quad . \quad . \quad \text{(III.)}$$

which forms the basis of the overall energy balance sheet for the plant in accordance with the following energy or power stream diagram expressed in, say, heat units per second :


[Potential heat in fuel-oil+Electrical energy for pumps]
$[P_i+(e_0+e_w)]$
Heat transformed into available mechanical energy
[B]
Gain in total heat and kinetic energy of working agent
$(H_0-H_i)+(K_0-K_i)$
Heat gained by surrounding air
$[\Sigma r]$
Heat gained by cooling water
$[(J_0+w_0)-W_i]$
Potential heat in unburnt fuel
$[P_0]$

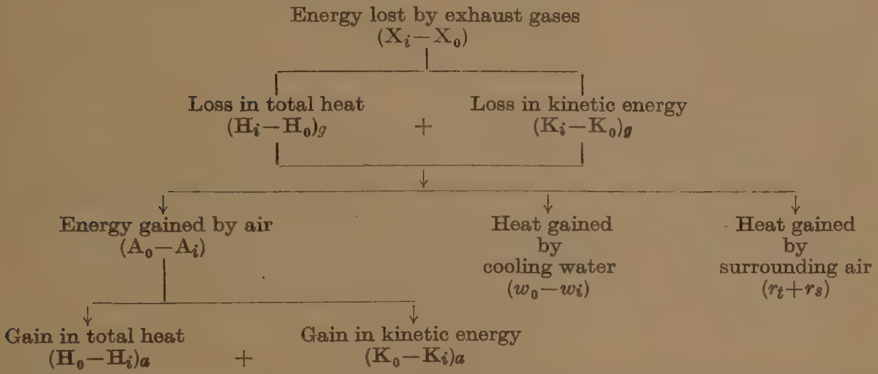
The "boundary method" indicated above can not only be applied to the whole plant, but also to any part of the plant. For instance, if it is desired to analyse the performance of the exhaust gas turbine and supercharger together as one unit, then from the inner boundary which encloses them both can be deduced the energy or power equation :

$$X_i+w_i+A_i=A_0+X_0+w_0+(r_t+r_s). \quad . \quad . \quad . \quad . \quad \text{(IV.)}$$

Assuming now that there is no further combustion while the working agent is flowing through the exhaust gas turbine, the energy of the gases entering and leaving the boundary by way of the inlet and outlet ducts will be composed of kinetic energy

$$\left(K=\frac{m.U^2}{2gJ}\right) \text{ and total heat } \left(H=E+\frac{p.v}{J}\right).$$

Thus $(X_i-X_0)=(H_i-H_0)+(K_i-K_0)$; and a similar equation represents the gain in energy of the induction air (A_0-A_i) during its passage through the supercharger. Hence, on making these substitutions in equation (IV.) we get an equation which forms the basis of the following energy or power-stream diagram (heat units per second) for the turbine and supercharger when considered as one unit :


Energy lost by exhaust gases
$(X_i - X_0)$
Loss in total heat
$(H_i - H_0)_g$
+
Loss in kinetic energy
$(K_i - K_0)_g$
Energy gained by air
$(A_0 - A_i)$
Heat gained by cooling water
$(w_0 - w_i)$
Heat gained by surrounding air
$(r_t + r_s)$
Gain in total heat
$(H_0 - H_i)_a$
+
Gain in kinetic energy
$(K_0 - K_i)_a$

The numerical values of the items of this power-stream diagram can be estimated from the formulae :

$$(H_i - H_0)_g = m . C_p . (T_i - T_0),$$

and

$$(K_i - K_0)_g = \frac{R . m^3}{2gJ}\left[\left(\frac{T_i}{p_i . a_i}\right)^2 - \left(\frac{T_0}{p_0 . a_0}\right)^2\right],$$

where m is the mass of exhaust gases flowing per second along the turbine inlet and outlet ducts of cross-sectional areas a_i and a_0, and absolute temperatures and pressures T_i, p_i and T_0, p_0 respectively : C_p is the mean specific heat of the exhaust gases at constant pressure, and R the exhaust gas constant.

Similar expressions apply to the air in the supercharger ducts, from which the numerical values of the energy gained by the induction air $(A_0 - A_i)$ can be found.

A power or energy balance sheet for the exhaust gas turbine and supercharger when considered together as one unit can therefore be drawn up in accordance with equation (IV.) and the power-stream diagram.

Thus has been indicated how a complete analysis can be made of the rates of flow of energy which travel along the interconnected streams of any power plant. And since modern power plant is becoming increasingly complex—especially in the larger electric generating stations, it seems desirable that a logical and rational method should be developed for the comprehensive analysis of such plant, and particularly one which obviates the possibility of erroneous items appearing on the energy balance sheets of individual units or of combinations of units of complex power plant.

The Northampton Polytechnic,
St. John Street,
London, E.C. 1.

LVIII. *Some Impedance Characteristics of Tapped Resonant-line Circuits.*

By W. E. Willshaw, M.Sc.Tech. (Communication from the Staff of the Research Laboratories of The General Electric Company Limited, Wembley, England*.)

[Received May 10, 1939.]

Summary.

The paper discusses some properties of tapped uniform lines relevant to their use as frequency stabilizing elements in short wave oscillators.

The *resonant impedance* of a tapped quarter-wave line, which has one end shorted, is given by $\frac{Z_0}{\alpha_0 L}\sin^2\beta_0 L_1$, where $\frac{Z_0}{\alpha_0 L}$ is the resonant impedance measured at the open end and $\beta_0 L_1$ is the length in electrical degrees between the tapping point and the closed end. L_1 is the physical length between the latter two points and L is the total line length, whilst α_0 and $\beta_0\left(=\frac{2\pi}{\lambda_0}\right)$ are respectively real and imaginary parts of the line propagation constant (=P) at the resonant frequency f_0. *The phase angle off resonance* is given by $-\tan^{-1}\frac{\delta\beta}{\alpha_0}$, where $\delta\beta$ is the change in β from its resonant value β_0. For small frequency changes from resonance (*i. e.*, so long as the line is performing its stabilizing function) the equivalent circuit is a substantially constant resistance shunted by a variable reactance.

The *resonant frequency* is independent of the position of the tapping point for all useful values of resonant impedance.

Experimental values for the input resistance of a concentric line at about 75 Mc/s are in good agreement with the calculated values over a range of positions of the tapping point.

Introduction.

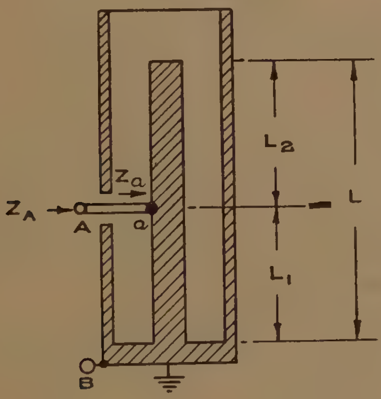
In the use of a low loss concentric line circuit to control the frequency of a short wave oscillator the arrangement shown in fig. 1 is convenient. Here a section of concentric line has an inner conductor approximately one-quarter wave-length long (or in some circumstances an odd number of quarter wave-lengths long), the inner conductor being shorted to the outer at one end. The outer is at earth potential and forms one of the connexions to the system, and a second connexion is made to a tapping point on the inner. By adjustment of the position of this tapping point the input impedance at terminals A and B may be adjusted to suit the circuit to which these two terminals are connected. The general requirements, as far as frequency control is concerned, are that at the resonant frequency of the system the input conductance across terminals AB should be as low as possible, whilst the susceptance for a given departure of frequency from resonance should be as high as possible.

---

* Communicated by Dr. C. C. Paterson.

Although the similarity with the corresponding lumped circuit case is obvious, it is not quite so clear how the resonant conductance and the susceptance off resonance vary with position of the point of connexion of terminal A to the inner conductor.

Fig. 1.



The object of this paper is to calculate these values as functions of tapping-point position, to present a simple equivalent circuit valid for frequencies near resonance, and to give some details of experimental checks made on the conductance near resonance of a circuit of the above type *.

The Admittance of Point " a " to Earth $(Y)_a$.

Calculations are considerably facilitated by working in admittances. The admittance Y_a between point " *a* " and earth is made up of two parts :

---

* A paper by M. Reed (' Wireless Engineer,' xv. p. 414), published during the preparation of the present one, includes an analysis of part of this problem. It is felt, however, that in view of the difference in method of treatment the interest of the present paper is not in any way reduced.

Y_1 due to the length of line L_1 short-circuited at its far end, and Y_2 due to the length of line L_2 open-circuited at its far end.

If the line constants are

Z_0=characteristic impedance,

$P=\alpha+j\beta$=propagation constant, then, using the ordinary transmission line equations

$$Y_a=Y_1+Y_2=\frac{1}{Z_0}(\coth PL_1+\tanh PL_2),$$

$\coth PL_1$ and $\tanh PL_2$ can be divided into real and imaginary parts by using the device

$$R(Y)=\tfrac{1}{2}(Y+\overline{Y}),$$

$$I(Y)=\tfrac{1}{2}(Y-\overline{Y}),$$

where $R(Y)$ and $I(Y)$ are the real and imaginary parts of Y, respectively, and $\overline{Y}$ is the conjugate of Y.

This gives

$$R(\coth PL_1)=\tfrac{1}{2}(\coth PL_1+\coth \overline{PL_1})$$

(since $\overline{\coth PL_1}=\coth \overline{PL_1}$)

$$=\tfrac{1}{2}\left[\frac{\cosh PL_1 \sinh \overline{PL_1}+\sinh PL_1 \cosh \overline{PL_1}}{\sinh PL_1 \sinh \overline{PL_1}}\right]$$

$$=\frac{\sinh (PL_1+\overline{PL_1})}{\cosh (PL_1+\overline{PL_1})-\cosh (PL_1-\overline{PL_1})}$$

$$=\frac{\sinh 2\alpha L_1}{\cosh 2\alpha L_1-\cos 2\beta L_1} \tag{1}$$

(substituting $\alpha+j\beta$ for P).

Similarly

$$I(\coth PL_1)=\frac{-\sin 2\beta L_1}{\cosh 2\alpha L_1-\cos 2\beta L_1}, \tag{2}$$

and

$$R(\tanh PL_2)=\frac{\sinh 2\alpha L_2}{\cosh 2\alpha L_2+\cos 2\beta L_2}, \tag{3}$$

$$I(\tanh PL_2)=\frac{\sin 2\beta L_2}{\cosh 2\alpha L_2+\cos 2\beta L_2}. \tag{4}$$

1. *Examination of* $I(Y_a)$.

$$Z_0 I(Y_a)=Z_0 I(Y_1+Y_2)$$

$$=-\frac{\sin 2\beta L_1}{\cosh 2\alpha L_1-\cos 2\beta L_1}+\frac{\sin 2\beta L_2}{\cosh 2\alpha L_2+\cos 2\beta L_2}$$

$$=-\frac{\sin 2\beta L_1}{(1-\cos 2\beta L_1)+(\cosh 2\alpha L_1-1)}+\frac{\sin 2\beta L_2}{(1+\cos 2\beta L_2)+(\cosh 2\alpha L_2-1)}. \qquad (5)$$

$$=-\frac{2\sin\beta L_1\cos\beta L_1}{2\sin^2\beta L_1+2\sinh^2\alpha L_1}+\frac{2\sin\beta L_2\cos\beta L_2}{2\cos^2\beta L_2+2\sinh^2\alpha L_2}$$

$$=-\frac{\cot\beta L_1}{1+\dfrac{\sinh^2\alpha L_1}{\sin^2\beta L_1}}+\frac{\tan\beta L_2}{1+\dfrac{\sinh^2\alpha L_2}{\cos^2\beta L_2}}. \qquad (6)$$

When $\dfrac{\sinh^2\alpha L_1}{\sin^2\beta L_1}$ and $\dfrac{\sinh^2\alpha L_2}{\cos^2\beta L_2}$ are $\ll 1$

$$Z_0 I(Y_a)=-\cot\beta L_1+\tan\beta L_2. \qquad (7)$$

The case where $\sin\beta L_1\rightarrow\sinh\alpha L_1$ and $\cos\beta L_2\rightarrow\sinh\alpha L_2$ does not enter into the present problem, since under these conditions $R(Y_a)$ at resonance is much too high for the circuit to be connected across any medium impedance generator.

For completeness and general information values of $I(Y_1)$ and $I(Y_2)$, when

$$\frac{\sinh^2\alpha L_1}{\sin^2\beta L_1} \text{ and } \frac{\sinh^2\alpha L_2}{\cos^2\beta L_2}\rightarrow 1 \text{ or are } >1,$$

are, however, treated separately in an Appendix. Fig. 2 shows $Z_0 I(Y_1)$ and $Z_0 I(Y_2)$ sketched out as functions of βL_1 and βL_2 respectively, the chief properties of $Z_0 I(Y_1)$ and $Z_0 I(Y_2)$ being indicated on the diagram.

2. *Examination of* $R(Y_a)$.

From (1), (3), and (5)

$$Z_0 R(Y_a)=\frac{\sinh 2\alpha L_1}{\cosh 2\alpha L_1-\cos 2\beta L_1}+\frac{\sinh 2\alpha L_2}{\cosh 2\alpha L_2+\cos 2\beta L_2}$$

$$=\frac{\sinh 2\alpha L_1}{2(\sin^2\beta L_1+\sinh^2\alpha L_1)}+\frac{\sinh 2\alpha L_2}{2(\cos^2\beta L_2+\sinh^2\alpha L_2)}. \qquad (8)$$

Neglecting $\sinh^2\alpha L_1$ and $\sinh^2\alpha L_2$ in comparison with $\sin^2\beta L_1$ and $\cos^2\beta L_2$ respectively, and putting $\sinh 2\alpha L_1=2\alpha L_1$ and $\sinh 2\alpha L_2=2\alpha L_2$,

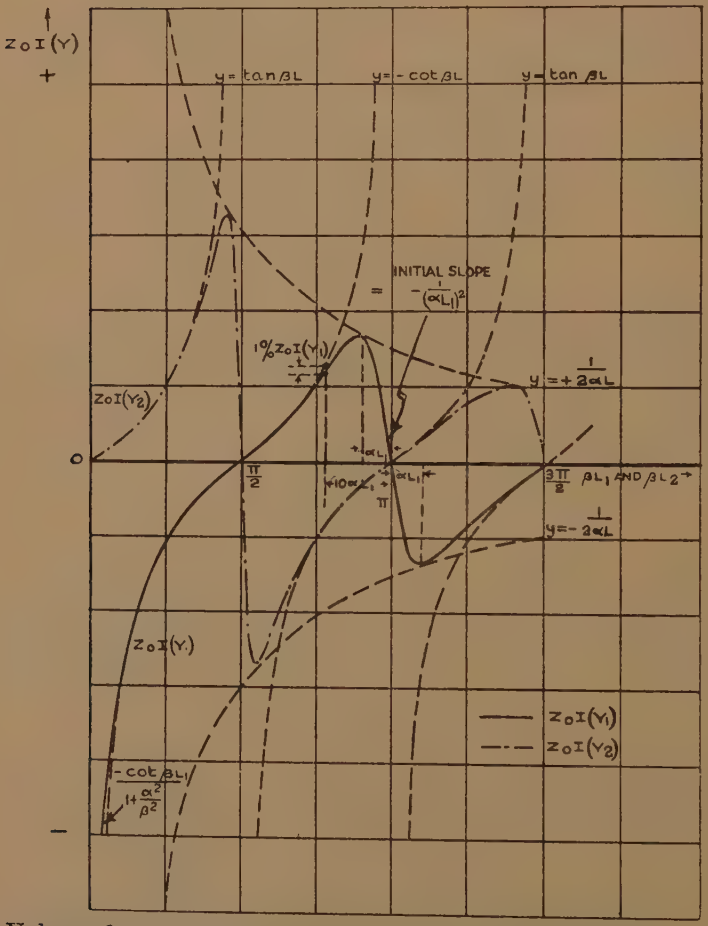
$$Z_0 R(Y_a)=\frac{\alpha L_1}{\sin^2\beta L_1}+\frac{\alpha L_2}{\cos^2\beta L_2}. \qquad (9)$$

Values of $Z_0R(Y_1)$ and $Z_0R(Y_2)$ for

$$\frac{\sinh \alpha L_1}{\sin \beta L_1} \rightarrow 1 \quad \text{and} \quad \frac{\sinh \alpha L_2}{\cos \beta L_2} \rightarrow 1$$

are again treated separately in the Appendix, and are sketched out in fig. 3.

Fig. 2.



Values of $Z_0I(Y_1)$ and $Z_0I(Y_2)$ as functions of βL_1 and βL_2.

Conditions for Resonance of Y_a.

Let resonant values of α, β, f, λ, etc. be denoted by α_0, β_0, f_0, λ_0, etc. (f being the frequency in cycles and λ the wave-length in cm.).

From (7) $\qquad Z_0I(Y_a)=(-\cot \beta L_1+\tan \beta L_2)$.

Resonance occurs when $Z_0I(Y_a)=0$ or when

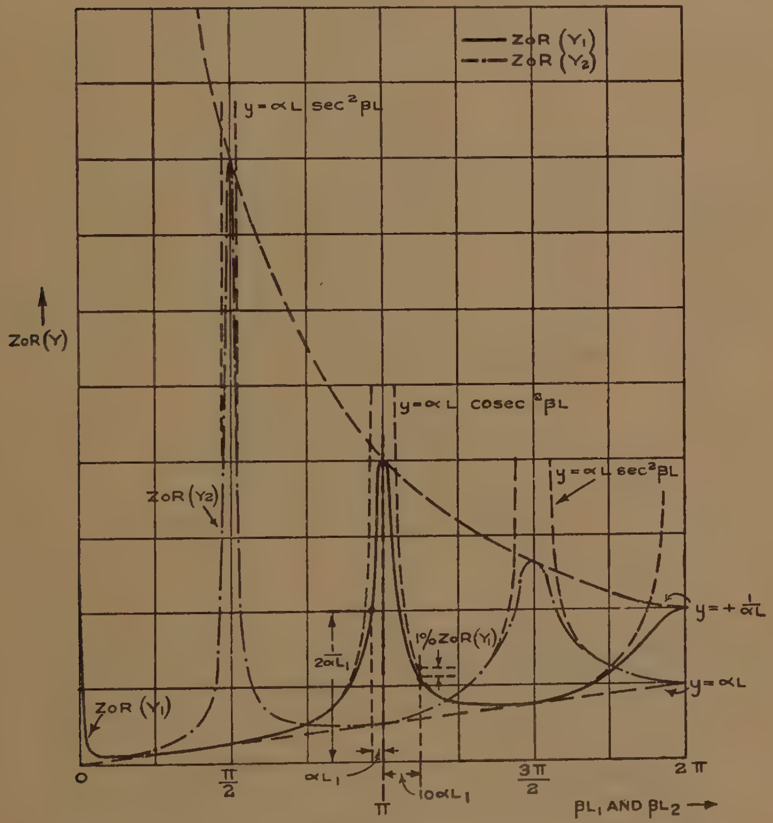
$$\beta_0L_2=n\pi+\left(\frac{\pi}{2}-\beta_0L_1\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

For the most usual case of $n=0$

$$\beta_0(L_1+L_2)=\frac{\pi}{2} \quad \text{or} \quad L_1+L_2=\frac{\lambda_0}{4}$$

$$\text{(substituting } \frac{2\pi}{\lambda_0} \text{ for } \beta_0\text{)}.$$

Fig. 3.



Values of $Z_0R(Y_1)$ and $Z_0R(Y_2)$ as functions of βL_1 and βL_2.

From this latter expression it is evident that the resonant wave-length λ_0 is dependent only on the total length L_1+L_2 so long as the condition

$$\frac{\sinh^2 \alpha L_1}{\sin^2 \beta L_1} \quad \text{and} \quad \frac{\sinh^2 \alpha L_2}{\cos^2 \beta L_1} \ll 1$$

is satisfied.

Phase Change of Y_a *about Resonance.*

The rate of change of $Z_0I(Y_a)$ about $Z_0I(Y_a)=0$ is given by

$$Z_0\left|\frac{d.I(Y_a)}{df}\right|_0=\left(\frac{d.\beta L_1}{df}\operatorname{cosec}^2\beta_0L_1+\frac{d.\beta L_2}{df}\sec^2\beta_0L_2\right)\quad . \quad (11)$$

$$\doteq(L_1+L_2)\operatorname{cosec}^2\beta_0L_1\,.\,\frac{d\beta}{df}\,.$$

And, for a small change of β $(=\delta\beta)$, from its resonant value β_0,

$$Z_0\delta I(Y_a)=(L_1+L_2)\operatorname{cosec}^2\beta_0L_1.\delta\beta\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$=(L_1+L_2)\operatorname{cosec}^2\beta_0L_1\,.\,\frac{2\pi}{\lambda_0}\,.\,\frac{\delta f}{f_0}\,.\quad . \quad . \quad . \quad . \quad . \quad (13)$$

From (9)

$$Z_0R(Y_a)=\frac{\alpha L_1}{\sin^2\beta L_1}+\frac{\alpha L_2}{\cos^2\beta L_2}\,.$$

And at resonance, putting in the condition of (10),

$$Z_0R(Y_{a0})=\alpha_0(L_1+L_2)\operatorname{cosec}^2\beta_0L_1.\quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Hence the phase angle for $\beta=\beta_0+\delta\beta$ is, from (12) and (14),

$$\theta_{Ya}=\tan^{-1}\frac{\delta I(Y)_a}{R(Y_a)}$$

$$=\tan^{-1}\frac{\delta\beta(L_1+L_2)\operatorname{cosec}^2\beta_0L_1}{\alpha_0(L_1+L_2)\operatorname{cosec}^2\beta_0L_1}$$

$$=\tan^{-1}\frac{\delta\beta}{\alpha_0}\,.\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The Resonant Impedance Z_{ao}.

If a and b are respectively the outer radius of the inner conductor and the inner radius of the outer conductor, in cm.,

$$Z_0=138\log_{10}.\frac{b}{a}\,.$$

Also $\alpha=\dfrac{r}{2Z_0}$,

where r=resistance per unit length of the inner conductor.

$=4{\cdot}15\times10^{-8}\sqrt{f}\left(\frac{1}{a}+\frac{1}{b}\right)$ ohm/cm. for copper.

and f=frequency in cycles/sec.

The resonant impedance Z_{ao} of the tapping point "a" to earth is then, from (14),

$$Z_{ao}=\frac{1}{R(Y_{ao})}=\frac{Z_0}{\alpha_0(L_1+L_2)}\sin^2\frac{2\pi L_1}{\lambda_0}$$

$$=\frac{2Z_0^2}{r_0(L_1+L_2)}\sin^2\frac{2\pi L_1}{\lambda_0},$$

and the phase angle off resonance is

$$\theta_{Z_a}=-\theta_{Y_a}=-\tan^{-1}\frac{\alpha\beta}{\alpha_0}=-\tan^{-1}\frac{\frac{2\pi}{\lambda_0}\cdot\frac{\delta f}{f_0}}{\alpha_0}. \quad . \quad . \quad . \quad . \qquad (16)$$

(The range of $\delta\beta$ over which this holds is discussed in the next section.)

The Equivalent Circuit of Z_a near Resonance.

The formulæ given above for θ are strictly only true for infinitely small values of $\delta\beta$. In practice it is necessary to consider how θ and $R(Y_a)$ (preferably) vary with finite values of $\delta\beta$. The most usual case of the *quarter wave-length* line is treated below.

From (9)

$$Z_0R(Y_a)=\frac{\alpha L_1}{\sin^2\beta L_1}+\frac{\alpha L_2}{\cos^2\beta L_2}$$

$$=\frac{\alpha_0 L_1\sqrt{1+\frac{\delta f}{f_0}}}{\sin^2\left[\beta_0 L_1\left(1+\frac{\delta f}{f_0}\right)\right]}+\frac{\alpha_0 L_2\sqrt{1+\frac{\delta f}{f_0}}}{\cos^2\left[\beta_0 L_2\left(1+\frac{\delta f}{f_0}\right)\right]}$$

(since α is proportional to $\sqrt{f}$).

Let $$\frac{\sin\left[\beta_0 L_1\left(1+\frac{\delta f}{f_0}\right)\right]}{\sin\beta_0 L_1}=1+p,\text{ where } 0<p<\frac{\delta f}{f_0},$$

and $$\frac{\cos\left[\beta_0 L_2\left(1+\frac{\delta f}{f_0}\right)\right]}{\cos\beta_0 L_2}=1-q,\text{ where } 0<q<\frac{\delta f}{f_0};$$

then $$Z_0R(Y_a)=\sqrt{1+\frac{\delta f}{f_0}}\left[\frac{\alpha_0 L_1}{(1+p)^2\sin^2\beta_0 L_1}+\frac{\alpha_0 L_2}{(1-q)^2\cos^2\beta_0 L_2}\right]. \qquad (17)$$

Further, putting $\beta=\beta_0$ in (9),

$$Z_0R(Y_{ao})=\frac{\alpha_0 L_1}{\sin^2\beta_0 L_1}+\frac{\alpha_0 L_2}{\cos^2\beta_0 L_2},$$

so $\frac{Z_0R(Y_a)}{Z_0R(Y_{ao})}$ must lie between $\frac{\sqrt{1+\frac{\delta f}{f_0}}}{(1+p)^2}$ and $\frac{\sqrt{1+\frac{\delta f}{f_0}}}{(1-q)^2}$,

depending on whether $\sin\beta_0L_1\to 1$ or $\cos\beta_0L_2\to 1$.

If $\frac{\delta f}{f_0}<0{\cdot}1$ say, then, since p and q are $<\frac{\delta f}{f_0}$,

$$\frac{Z_0R(Y_a)}{Z_0R(Y_{ao})} \text{ must lie between } \left(1+\frac{\delta f}{f_0}\right)^{-3/2} \text{ and } \frac{\left(1+\frac{\delta f}{f_0}\right)^{1/2}}{\left(1-\frac{\delta f}{f_0}\right)^2},$$

or between $\left(1-\frac{3}{2}\frac{\delta f}{f_0}\right)$ and $\left(1+\frac{5}{2}\frac{\delta f}{f_0}\right)$ very nearly.

Hence $\frac{Z_0R(Y_a)}{Z_0R(Y_{ao})}-1$ is certainly less than $\pm\frac{5}{2}\frac{\delta f}{f_0}$ for $\frac{\delta f}{f_0}<0{\cdot}1$. . (18)

Further, from (11)

$$Z_0\frac{dI(Y_a)}{df}=\frac{d\beta}{df}\left(\frac{L_1}{\sin^2\beta L_1}+\frac{L_2}{\cos^2\beta L_2}\right),$$

and similar treatment indicates that

$$\frac{Z_0\frac{dI(Y_a)}{df}}{Z_0\left|\frac{dI(Y_a)}{df}\right|_0}-1$$

is certainly less than

$$\pm 2\frac{\delta f}{f_0} \text{ for } \frac{\delta f}{f_0}<0{\cdot}1. \quad \ldots\ldots\ldots \quad (19)$$

The conclusions contained in (18) and (19) show that variations of $Z_0R(Y_a)$ and $Z_0\frac{dI(Y_a)}{df}$ from their resonant values are certainly less than $\pm 2\frac{1}{2}$ and ± 2 per cent. respectively for a frequency change $\frac{\delta f}{f_0}$ of 1 per cent.

Further, since changes of $Z_0R(Y_a)$ and $Z_0\frac{dI(Y_a)}{df}$ are in the same direction, the difference between

$$\frac{I(Y_a)}{R(Y_a)}=\frac{\int_0^{\delta f}\frac{dI(Y_a)}{df}}{R(Y_a)}.df$$

and the approximate value

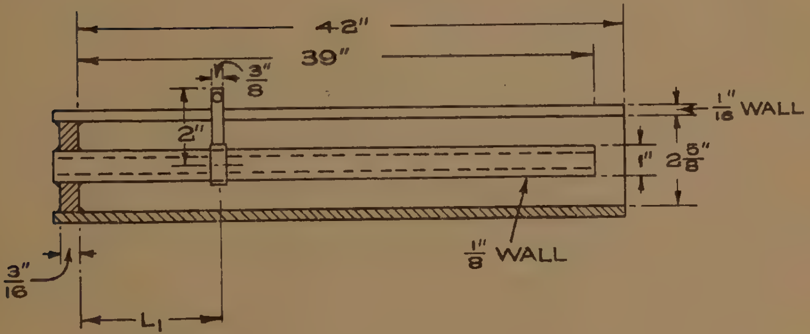
$$\frac{\left|\frac{dI(Y_a)}{df}\right|_0 . \delta f}{R(Y_{ao})} = \frac{\delta\beta}{\alpha_0} \quad . \quad . \quad ((12), (14), \text{and } (15))$$

is less than $2\frac{1}{2}$ per cent. for $\frac{\delta f}{f_0}=1$ per cent.

Considering now the tube circuit whose dimensions are given in fig. 4. At a resonant frequency of 75 Mc/s.

$r_0=3{\cdot}91\times10^{-4}$ ohm/cm.
$Z_0=58$ ohm.
$\alpha_0=3{\cdot}37\times10^{-6}$ nepers/cm.
$\beta_0=\frac{2\pi}{\lambda_0}=1{\cdot}57\times10^{-2}$ radians/cm.

Fig. 4.



For $\frac{\delta f}{f_0}=1$ *per cent.*

$$\delta\beta=1{\cdot}57\times10^{-4}.$$

And $$\theta_{Ya} \doteq \tan^{-1}\frac{\delta\beta}{\alpha_0}=\tan^{-1}\frac{1{\cdot}57\times10^{-4}}{3{\cdot}37\times10^{-6}}.$$

Hence $$\theta_{Ya} \doteq 88^\circ{\cdot}77 \text{ for } \frac{\delta f}{f_0}=1 \text{ per cent.}$$

This result and the conclusions to (18) and (19) show that in the particular example chosen the variation of $Z_0R(Y_a)$ from its resonant value is less than $2\frac{1}{2}$ per cent. when θ_{Ya} is nearly 88°·8.

It may be taken as a general rule that the equivalent circuit of Z_a near resonance is the constant resistance $\frac{Z_0}{\alpha_0(L_1+L_2)}\sin^2\frac{2\pi L_1}{\lambda_0}$ shunted by a variable reactance of such a value as to give Z_a a phase angle $\theta_{za}=-\tan^{-1}\frac{\delta\beta}{\alpha_0}$. This is accurate to well within 1·25 per cent. in magnitude and angle if $\frac{\delta f}{f_0}$ is not greater than 0·5 per cent.

Z_a may then be represented by the circle diagram shown in fig. 5, in which

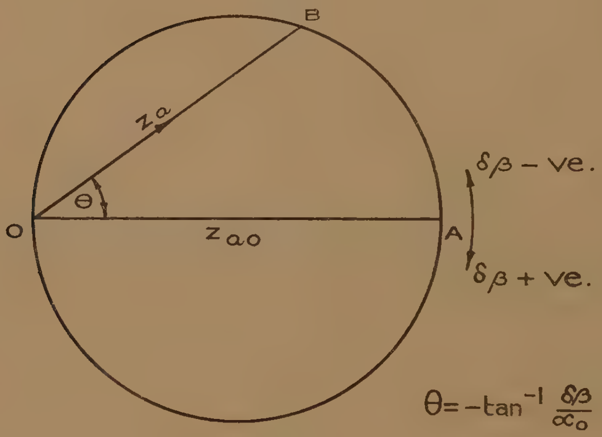
$$OA=Z_{ao}=\frac{2Z_0^2}{r_0(L_1+L_2)}\sin^2\frac{2\pi L_1}{\lambda_0}$$

and

$$OB=Z_a$$

$$\angle_{AOB}=\theta=-\tan^{-1}\frac{\delta\beta}{\alpha_0}.$$

Fig. 5.



Experimental Checks.

Tests were made on the input resistance near resonance of the circuit already quoted for different positions of the tapping point. The frequency was adjusted so that the terminal impedance Z_A was resistive (see fig. 1), the impedance Z_a being off resonance by an amount necessary to offset the effect of the capacity to earth of the tapping conductor. This meant that for values of L_1 of approximately 1/8 (L_1+L_2) the measuring frequency was close to the resonant frequency of the tube, whilst for $L_2=0$ it differed by —5 per cent., corresponding to a capacity to earth of 2·9 $\mu\mu$F. for the tapping conductor.

Two sets of readings were taken, one with a shorting connexion between inner and outer conductors formed by a thin copper disk (approximately 1/16 in. thick), provided with four lugs ½ in. wide, round its circumference. This disk was connected to the inner conductor by means of a split sleeve soldered on to the disk, and clamped on to the conductor. Connexion to the outer tube was made by clamping down the lugs with a brass band, $\frac{1}{2}$ in. wide.

The second set of readings was taken with a 3/16 in. copper disk soldered up to inner and outer conductors to form a really solid joint.

Details of dimensions are given in fig. 4. All material was copper except for clamping bands.

In the table of results given below the "calculated" impedance includes an allowance for the effect of frequency change on α and β, the impedance being calculated from the value given by equation (9).

Test Results with Soldered End Connexion.

L_1 (cm.).	12·5.	25.	50.	95.
Frequency (Mc/s.) .	74·55	74·25	72·9	71·1
Measured Z_A (ohm) .	6,400	24,400	79,000	159,000
Calculated Z_A (ohm).	6,650	25,500	89,000	178,000

Measured d.c. resistance $1{\cdot}32\times10^{-4}$ ohm.
Calculated d.c. resistance $1{\cdot}32\times10^{-4}$ ohm.

Test Results with Clamped End Connexion.

L_1 (cm.).	12·5.	25.	50.	95.
Measured Z_A (ohm).	4,530	14,400	57,000	115,000

Measured d.c. resistance $1{\cdot}85\times10^{-4}$ ohm.

The d.c. resistance values given were measured from "free" end of inner conductor to "free" end of outer conductor. The calculated value of d.c. resistance is on the assumption of $\rho=1{\cdot}72\times10^{-6}$ ohm/cm.^{3}, for which value the formula for h.f. resistance already quoted holds.

Comments on Test Results.

The above tables show reasonable agreement between measured and calculated values of Z_A with the soldered end connexion. The apparatus on which the impedances were measured was actually more suited for measuring resistances of about 10,000 ohm than of about 100,000 ohm. Consequently the measurements for $L_1=12{\cdot}5$ cm. and $L_1=25$ cm. are the more reliable, and agree with calculated values to within 5 per cent.

With the clamped end connexion measured values of Z_A were only about two-thirds of those for the solid connexion. The defective joint is shown up to some extent in the d.c. resistance figures. It is evident that if full theoretical impedances are to be obtained the end connexion must be a good one.

Conclusion.

It is apparent that a line circuit of the type discussed becomes more useful in its function as a sharply resonant impedance as its resonant frequency is increased. For fixed diameters of inner and outer conductors, and given fractional change of frequency from resonance, $\tan\theta_{za}$ is proportional to $\sqrt{f_0}$ ($\delta\beta$ is proportional to f_0 and α_0 is proportional to $\sqrt{f_0}$). The minimum length of connecting lead between any other circuit and the actual tapping point "*a*" may, however, become an appreciable fraction of a wave-length at higher frequencies. Consequently it should be pointed out that unless special precautions are taken at higher frequencies the impedance Z_A measured at the end of the connecting lead remote from the tapping point "*a*" may be greatly different from that indicated by the above theory.

Appendix.

1. *Values of* $I(Y_1)$ *and* $I(Y_2)$ *when* $\sin\beta L_1\rightarrow\sinh\alpha L_1$ *and* $\cos\beta L_2\rightarrow\sinh\alpha L_2$ (αL_1 and αL_2 small)

From (2) and (5)

$$Z_0 I(Y_1) = -\frac{\sin 2\beta L_1}{2\sin^2\beta L_1 + 2\sinh^2\alpha L_1}.$$

Putting $\sinh\alpha L_1 = \alpha L_1$ and $\beta L_1 = n_1\pi + \phi_1$,

where n_1 is an integer, then, when $\phi_1 = \alpha L_1$,

$$\sin\beta L_1 \doteq \phi_1 \quad\text{and}\quad \sin 2\beta L_1 \doteq 2\phi_1.$$

Hence $$Z_0 I(Y_1) = -\frac{\phi_1}{\phi_1^2 + (\alpha L_1)^2}.$$

From this it is evident that $Z_0 I(Y_1)$ has a maximum value of $-\frac{1}{2\alpha L_1}$ at $\phi_1 = \alpha L_1$, is 0 at $\phi_1 = 0$, and has a slope $\frac{dZ_0 I(Y_1)}{d\phi_1}$ of $-\frac{1}{(\alpha L_1)^2}$ at $\phi_1 = 0$. In addition when $\phi_1 = 10\alpha L_1$, $Z_0 I(Y_1)$ differs from $-\cot\phi(\doteq -1/\phi)$ by only 1 per cent. (See fig. 2.)

For the particular case of $L_1\rightarrow 0$; $n_1 = 0$ and $\phi_1 = \beta L_1$,

hence $$Z_0 I(Y_1) = -\frac{\cot\beta L_1}{1+\frac{\alpha^2}{\beta^2}} \quad \text{. . . . . . . (From (6))}$$

Similar manipulation of $Z_0I(Y_2)$ (4) and (5)) for $\beta L_2=n_2\pi+\frac{\pi}{2}+\phi_2$, where $\phi_2\rightarrow\alpha L_2$, yields similar results, and values of $Z_0I(Y_2)$ are also indicated in fig. 2.

2. *Values of* $R(Y_1)$ *and* $R(Y_2)$ *when* $\sin\beta L_1\rightarrow\sinh\alpha L_1$ *and* $\cos\beta L_2\rightarrow\sinh\alpha L_2$ (αL_1 and αL_2 small).

From (8)

$$Z_0R(Y_1)=\frac{\sinh 2\alpha L_1}{2\sin^2\beta L_1+2\sinh^2\alpha L_1}.$$

Again putting $\sinh\alpha L_1=\alpha L_1$, the following relationships hold :—

When

$\sin\beta L_1=0,\quad Z_0R(Y_1)=\frac{1}{\alpha L_1}$,

$\sin\beta L_1=\alpha L_1,\quad Z_0R(Y_1)=\frac{1}{2\alpha L_1}$,

$\sin\beta L_1=10\alpha L_1,\quad Z_0R(Y_1)=\alpha L_1\operatorname{cosec}^2\beta L_1$, to an accuracy of 1 per cent.

$\sin\beta L_1=1,\quad Z_0R(Y_1)=\alpha L_1$.

Similar relations hold if L_1, $\sin\beta L_1$, $R(Y_1)$ are replaced by L_2, $\cos\beta L_2$, and $R(Y_2)$ respectively.

Fig. 3 shows values of $Z_0R(Y_1)$ and $Z_0R(Y_2)$ which are seen to lie between the maximum and minimum values of $1/\alpha L$ and αL.

In conclusion, it should be pointed out that figs. 2 and 3 are of necessity distorted, and correspond to values of αL much larger than would justify the approximation $\left(\frac{\sinh\alpha L}{\alpha L}\doteqdot 1\right)$ upon which they are based.

Research Laboratories,
General Electric Company,
Wembley,
England.

LIX. *Differential Operators acting as Integrators.*

By Ludwik Silberstein, Ph.D.*

[Received April 20, 1939.]

Let $f=f(\mathrm{P})$ be a one-valued scalar function of position or scalar field, in a Euclidean manifold : space, plane, or line. Let $\bar{f}$ be the arithmetical mean of its values taken over the locus of all points Q equidistant from $\mathrm{P}(x, y, z)$, that is to say, over a sphere or circle of radius a and centre P, and over a pair of points $(x+a, x-a)$, in the three cases respectively. It is required to find the differential operator Ω which, applied to f at the centre P, yields this mean value,

$$\bar{f}=\Omega f.$$

We may call it an *integrator* or an *averaging operator*, viz., spherical or circular in the first two cases. The remarkable thing about this operator is that it will be seen to contain *only the Laplacean* ∇^2 and its iterations, the Laplacean being defined intrinsically, without the aid of a reference system, by

$$\nabla^2=\text{divergence of gradient}=\text{div}\,\nabla,$$

which holds alike for the space, the plane, and the line. (In the latter case, of course, $\nabla^2=d^2/dx^2$.)

Let $\mathbf{a}$ be the vector, of size a, drawn from P to any point Q of the locus. Let it be granted that the field f is, in general, developable into a Taylor series about any point P, with some radius of convergence, and let it be assumed throughout the discussion that a does not exceed this radius. Then the field at Q will be

$$f+(\mathbf{a}\nabla)f+\frac{1}{2\,!}(\mathbf{a}\nabla)^2f+\ldots.=\{e^{\mathbf{a}\nabla}\}f,$$

and $\bar{f}=\overline{\{e^{\mathbf{a}\nabla}\}}f$ or, the operator itself,

$$\Omega=\overline{e^{\mathbf{a}\nabla}},$$

to be averaged over all directions of the vector $\mathbf{a}$. It remains to perform this averaging process. Now, first of all, take the mean for any pair of antipodes ($\mathbf{a}$ and $-\mathbf{a}$). Their joint contribution is $\frac{1}{2}(e^{\mathbf{a}\nabla}+e^{-\mathbf{a}\nabla})=\cosh(\mathbf{a}\nabla)$, so that

$$\Omega=\overline{\cosh\,(\mathbf{a}\nabla)}, \quad \ldots\ldots\ldots \quad (1)$$

---

* Communicated by the Author.

henceforth to be averaged over half the locus (hemisphere, semicircle, single Q-point). Next, by the definition of the gradient,

$$\nabla f = \mathbf{n}\frac{\partial f}{\partial n},$$

or

$$\nabla = \mathbf{n}\frac{\partial}{\partial n},$$

where $\mathbf{n}$ is a unit vector in the direction of the most rapid increase of f at the centre P, which is thus a unique *fixed* direction *, and $\partial/\partial n$ the ordinary differentiator along this direction. Thus, if θ be the angle between the mobile vector $\mathbf{a}$ and the fixed $\mathbf{n}$,

$$\mathbf{a}\nabla = a\cos\theta \, . \, \frac{\partial}{\partial n} .$$

But, $\mathbf{n}$ being a unit vector, $\nabla^2 = \frac{\partial^2}{\partial n^2}$, and since the operator (1) contains only even " powers " of $(\mathbf{a}\nabla)$, and since, actually,

$$\overline{(\mathbf{a}\nabla)^2} = a^2\, \overline{\cos^2\theta} \, . \nabla^2, \quad \overline{(\mathbf{a}\nabla)^4} = a^4\, \overline{\cos^4\theta} \, . \nabla^4, \quad \text{and so on,}$$

it is legitimate † to write, in (1), $\mathbf{a}\nabla = a\cos\theta \, . \nabla$ or $a\nabla \, . \cos\theta$.

Thus our operator becomes, for all three cases,

$$\Omega = \overline{\cosh\left(a\nabla \, . \cos\theta\right)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

to be averaged over (half the field of) the angle θ. At the same time we see that it contains *only the Laplacean* ∇^2, and its iterations. It is thus an intrinsic operator, of scalar character, and distributive, as well as ∇^2 itself, $\Omega(f+g) = \Omega f + \Omega g$. If the operand f happens to be a harmonic function or harmonic field ($\nabla^2 f = 0$), then $\Omega = 1$, an idemfactor, that is to say, $\bar{f} = f$, for any legitimate radius a, a most characteristic property of all such fields, well known from potential theory and the theory of functions.

---

* We discard, of course, the trivial case of radial symmetry about P, when $\mathbf{n}$ is *any* radial vector (and $\theta = 0$ for every $\mathbf{a}$). In this case, however, no averaging is needed, f being constant on the circle or the sphere.

† The legitimacy of this short cut may be verified by Cartesian expansion. Thus, *e. g.*, for two dimensions, with $\mathbf{a} = a\,(\mathbf{i}\cos\alpha + \mathbf{j}\sin\alpha)$,

$$\overline{(\mathbf{a}\nabla)^2} = a^2\left(\overline{\cos^2\alpha}\,\frac{\partial^2}{\partial x^2} + 2\,\overline{\sin\alpha\cos\alpha}\,\frac{\partial^2}{\partial x\,\partial y} + \overline{\sin^2\alpha}\,\frac{\partial^2}{\partial y^2}\right) = a^2\,\overline{\cos^2\alpha} \, . \nabla^2 .$$

Similarly, if we write down the surviving terms only,

$$\overline{(\mathbf{a}\nabla)^4} = a^4\left[\overline{\cos^4\alpha}\left(\frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4}\right) + 6\,\overline{\sin^2\alpha\cos^2\alpha}\,\frac{\partial^4}{\partial x^2\,\partial y^2}\right] = a^4\,\overline{\cos^4\alpha} \, . \nabla^4,$$

and so on. And the same property holds in three dimensions.

It remains now to perform the averaging over θ indicated in the general expression (2), for each of the three cases, of one, two, and three dimensions, *e. g.*, for $f=f(x)$, $f(x, y)$, and $f(x, y, z)$.

1. LINE.—In this case there is no averaging to be done. For we have already paired together the two points (Q) equidistant from P, and since in this case **a** coincides with **n**, $\cos\theta=1$, and the operator is simply

$$\Omega=\cosh\,(a\nabla) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(I)}$$

or $\cosh\left(a\frac{d}{dx}\right)$. In developed form, and inserting the operand,

$$\bar{f}=f+\frac{a^2}{2\,!}\nabla^2 f+\frac{a^4}{4\,!}\nabla^4 f+\ldots \quad . \quad . \quad . \quad . \quad . \quad . \quad \text{(I } a\text{)}$$

Here, $\nabla^2 f=f''=\frac{d^2 f}{dx^2}$, but it is preferable, for the sake of uniformity and comparison, to retain the same symbol ∇^2 for all three cases. The mean is now, simply, $\bar{f}=\frac{1}{2}f(x+a)+\frac{1}{2}f(x-a)$.

If f is a "harmonic" function, $\nabla^2 f=f''=0$, the operator is an idemfactor or $\bar{f}=f$, for *any* "radius" a, which in the present case is the trivial property of all linear functions,

$$f=\alpha+\beta x.$$

Conversely, if we claim the property $\bar{f}=f$ or $\Omega=1$ for all a, the only solution of

$$\cosh\,(a\nabla)=1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\text{I}_0)$$

is $\nabla^2=0$, *i. e.*, a linear function f. But if the property is claimed only for some particular, given a, then (I_0), which is equivalent to a differential equation of infinitely high order, has an infinity of other solutions. Even these are trivial enough, but for the sake of comparison with the other two cases may deserve a moment's attention. The equation (I_0) is satisfied by $\nabla^2=m^2$, a constant number, *i. e.*, by $f=ce^{mx}$, provided that m satisfies the equation

$$\cosh\,(am)=1.$$

The only real root of this transcendental equation is $m=0$ (a double root), already considered. Apart from this, there is an infinity of conjugate imaginary pairs of roots, namely, $am=\pm 2\pi ni$, $n=1, 2, 3$, etc., and there are no other roots. The coefficients c of the exponentials corresponding to any of these roots remain perfectly arbitrary. Thus, and since (I_0) is a linear equation with constant coefficients, its complete solution is

$$f=\alpha+\beta x+\text{F}(x),$$

where F is any periodic function of period a. This, then, exhausts the class of functions endowed with the particular property $\bar{f}=f$, an obvious result.

2. PLANE.—The expression in (2) is, in this case, to be averaged over a semicircle, *i. e.*, over $0 \leqslant \theta \leqslant \pi$. We may as well write θ instead of $\frac{\pi}{2}-\theta$. Thus, and since $\cosh\phi=\cos(i\phi)$, our operator becomes

$$\Omega=\overline{\cos(ia\nabla\,.\,\sin\theta)}=\frac{1}{\pi}\int_0^\pi \cos(ia\nabla\,.\,\sin\theta)\,.\,d\theta.$$

But this is the zeroth Bessel function, J, with $ia\nabla$ as its argument.

Thus the averaging-operator for the plane or, in fact, the *circular integrator*, receives the compact expression

$$\Omega=\mathrm{J}(ia\nabla), \quad . \; . \; . \; . \; . \; . \; . \; . \; . \; . \quad \text{(II)}$$

where a is the radius of the circle. Needless to say, it is an *even* function of $i\nabla$, and thus real, and contains only the Laplacean ∇^2. In developed form, and inserting the field-operand,

$$\bar{f}=f+\frac{a^2}{2^2}\nabla^2 f+\frac{a^4}{2^2\,.\,4^2}\nabla^4 f+\frac{a^6}{2^2\,.\,4^2\,.\,6^2}\nabla^6 f+\;.\;.\;.\;.\quad . \quad \text{(II } a\text{)}$$

Whether, and how rapidly, this generally infinite series converges, depends on the properties of the contemplated field and the choice of the radius a. This question has no place in an investigation of the structure of the operator itself. If the series diverges and thus becomes meaningless, it is, so to speak, not the fault of the operator, but of the operand. There are, moreover, a good many cases of interest in which the series terminates and consists only of a few terms, or can otherwise be evaluated at once, as will be seen later on from a few examples. Again, differentiations are easily accomplished, and the utility of the operator Ω consists, then, therein that it enables us to evaluate integrals around the circle by differentiating the field at its centre.

In the one-dimensional case we have had $\Omega=\cos(ia\nabla)$, containing only $a^2\nabla^2$. In two dimensions the situation is perfectly analogous, only the cosine is replaced by the simplest Bessel function.

If $\nabla^2 f=0$, say, in Cartesians,

$$\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}=0,$$

then, as has already been mentioned, $\mathrm{J}=1$, Ω becomes an idemfactor and gives $\bar{f}=f$, thus evaluating an integral at a stroke. The most general function endowed with the property $\bar{f}=f$ is now the complete integral of the equation

$$\mathrm{J}(ia\nabla)=1, \quad . \; . \; . \; . \; . \; . \; . \; . \; . \quad (\mathrm{II}_0)$$

a linear differential equation with constant coefficients, of infinitely high order. If the property is required to hold generally, apart from singularities, for any radius a, the only solution of the equation is $\nabla^2 f=0$, any two-dimensional harmonic function, *e. g.*, $f=\log r$, where r is the distance from some fixed point O (outside the circle a). These, however, are only the simplest solutions of (II_0); they correspond to $f=\alpha+\beta x$ of the one-dimensional case, although they now seem less trivial, being represented by the vast class of two-dimensional harmonics. If the property $\bar{f}=f$ is claimed only for one value of a, at a time, the equation (II_0), or

$$1+\frac{1}{4^2}a^2\nabla^2+\frac{1}{4^2 . 6^2}a^4\nabla^4+\ldots=0,$$

has an infinite variety of other solutions, *e. g.*, solutions of the equation

$$\nabla^2 f=-m^2 f,$$

where the number m satisfies the equation $\mathrm{J}(am)=1$, which, besides $m=0$ (already covered by the harmonics), has an infinity of complex roots m_κ. Thus also $f=\Sigma c_\kappa f_\kappa$, with arbitrary coefficients c, will have the said property.

Returning to the general formula (II), we see that the first term of the central defect $\bar{f}-f$ is now $\frac{1}{2 . 2}a^2\nabla^2 f$, instead of the one-dimensional $\frac{1}{2}a^2\nabla^2 f$. We shall see presently that in three dimensions it is $\frac{1}{3 . 2}a^2\nabla^2 f$. Concrete examples, illustrating the use of the circular integrator (II), are given at the end of the paper.

3. SPACE.—The locus of points (Q) is in this case the sphere (surface) of radius a, centred at P, and the average indicated in the general expression (2) is to be taken over a hemisphere or solid angle 2π. Thus, after an obvious integration over the longitude (0 to π),

$$\Omega=\frac{\pi}{2\pi}\int_0^\pi \cosh\,(a\nabla . \cos\theta)\sin\theta\, d\theta=\frac{1}{2}\int_{-1}^{1}\cosh\,(a\nabla . x)\, dx,$$

so that the expression for *the spherical integrator* becomes

$$\Omega=\frac{\sinh\,(a\nabla)}{a\nabla}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{III})$$

The inverse, ∇^{-1}, occurs merely in appearance. The operator, of course, contains again only $a^2\nabla^2$ and its iterations. In developed form,

$$\bar{f}-f=\frac{1}{3\,!}a^2\nabla^2 f+\frac{1}{5\,!}a^4\nabla^4 f+\ldots\ldots \quad . \quad . \quad . \quad . \quad (\mathrm{III}\,a)$$

As mentioned a moment ago, the first term of the central defect is $\frac{1}{6}a^2\nabla^2 f$, perfectly analogous to that of the one-dimensional case *.

The simplest functions satisfying the particular requirement $\bar{f}=f$ are now again the solutions of $\nabla^2=0$, in fine, the three-dimensional harmonics. All such and only such functions have this property, in general, for any radius a. If the property $\bar{f}=f$ is claimed only for a given radius a, then the equation $\Omega=1$ or

$$\sinh(a\nabla)=a\nabla \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (\mathrm{III}_0)$$

has a vast variety of other solutions, *e. g.*, any solution of the differential equation

$$\nabla^2 f=m^2 f, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (3)$$

where the number m satisfies the transcendental equation

$$\sinh(am)=am,$$

which has again (apart from the only real root $m=0$ already considered) an infinity of complex roots, conjugate in pairs. In fact, if we put $am=u+iv$, the equation splits into the two equations

$$\sinh u\,.\cos v=u, \quad \cosh u\,.\sin v=v,$$

which are easily solved by successive approximations. There are no purely imaginary roots (since $u=0$ leads to $\sin v=v=0$), nor any roots within $0<v<2\pi$. All roots are of the form $v=2\pi n+\alpha$, where α is an acute angle. I may mention that the rapidly convergent process gives, as the lowest pair of roots, to four figures, $am=2{\cdot}768\pm i(2\pi+69^\circ{\cdot}54)$. The next pair is of the form $u\pm i(4\pi+\alpha)$, where α is about 80°, and so on, tending to $u\pm i(2\pi n+\pi/2)$. But details of this kind need not detain us here. To every root m_κ will correspond a whole class of functions

---

* It may be mentioned that Maxwell has considered the average of a scalar over *the volume* of a sphere. In vol. i. art. 26 of his Treatise, Maxwell states that for a "very small" sphere, in our notation, $\bar{f}-f=\frac{1}{10}a^2\Delta^2 f$.

In fact, our a^2-term, integrated over the volume, receives at once the numerical factor $\frac{1}{6}\cdot\frac{3}{5}=\frac{1}{10}$, in agreement with Maxwell's result. The complete operator (III) can be integrated over a with equal ease, giving for the *volume* average

$$\bar{f}_{\text{vol.}}=3\left\{\frac{a\nabla.\cosh(a\nabla)-\sinh(a\nabla)}{(a\nabla)^3}\right\}f=f+\frac{1}{10}a^2\nabla^2 f+\text{etc.}$$

It is thus enough to consider the surface average, which seems also more fundamental. Similarly, in two dimensions, the integral along the circumference, and not that over the area of the circle, is the fundamentally relevant concept, as will be familiar from the general theory of functions.

f_κ, and $f=\Sigma c_\kappa f_\kappa$, with arbitrary coefficients, will have the said property. Even if we limit ourselves to functions of the form $f=f(r)$, where r is the distance from a fixed point, when (3) reduces to an ordinary differential equation, this will give a great variety of fields for which Ω will become an idemfactor, thus evaluating a number of spherical surface integrals,

$$\bar{f}=\frac{1}{4\pi a^2}\int f\,d\sigma=f$$

at the centre.

The relevant point, if I may once more emphasize it, is that the wholly *differential* operators Ω actually, and with comparative ease, evaluate *integrals* extended over a sphere or a circle, and thus, apart from their conceptual interest, seem to offer some advantages. Illustrations will be given presently. But before passing to these, a few more general remarks about these operators may be inserted here.

Either operator, being of purely scalar character and distributive, may be applied at once to a *vector field* $\mathbf{R}$, yielding $\mathbf{R}=\Omega\mathbf{R}$, again a vector, the mean field, in size and direction, over the surface of a sphere or the circumference of a circle. Such means or integrals, however, are scarcely of any interest. What is of interest is, in the three-dimensional case, the integral of the normal component of $\mathbf{R}$ or the flux of the field across the surface, and in the two-dimensional case, the flux across the circumference of a circle, $\int \mathbf{aR}\,d\theta$, and the so-called line-integral around the circle which can be simply written $\int \mathbf{R}\,d\mathbf{a}$ (*e. g.*, the "circulation" in the case of liquid motion). It may suffice to treat in detail the two-dimensional field.

If $\mathbf{R}$ be the field at the centre of the circle, then the field at any point of its circumference is $\{e^{\mathbf{a}\nabla}\}\mathbf{R}$, and the flux is defined by

$$S=\int_0^{2\pi}\mathbf{a}\{e^{\mathbf{a}\nabla}\}\mathbf{R}.d\theta.$$

Let $\mathbf{i}$, $\mathbf{j}$ be a fixed pair of orthogonal unit vectors, $\mathbf{R}=\mathbf{i}R_1+\mathbf{j}R_2$, and $\mathbf{a}=a(\mathbf{i}\cos\theta+\mathbf{j}\sin\theta)$. Then

$$\frac{1}{2\pi a}S=\overline{\cos\theta\{e^{\mathbf{a}\nabla}\}}R_1+\overline{\sin\theta\{e^{\mathbf{a}\nabla}\}}R_2\,.$$

Now, if we write ∇_1, ∇_2 for $\partial/\partial x$, $\partial/\partial y$, the "components" of ∇,

$$\begin{aligned}\overline{\cos\theta\{e^{\mathbf{a}\nabla}\}}=\overline{\cos\theta}+a\,\overline{\cos\theta(\cos\theta.\nabla_1+\sin\theta.\nabla_2)}\\ +\frac{a^2}{2\,!}\overline{\cos\theta(\cos^2\theta.\nabla_1^2+2\sin\theta\cos\theta.\nabla_1\nabla_2+\sin^2\theta.\nabla_2^2)}\\ +\frac{a^3}{3\,!}\overline{\cos\theta(\cos^3\theta.\nabla_1^3+3\cos\theta.\sin^2\theta.\nabla_1\nabla_2^2)}\end{aligned}$$ [*] ....

---

[*] Omitting the terms which do not survive. None of the terms with a^2, a^4 etc. survive.

$$=\frac{a}{2}\nabla_1+\frac{a^3}{3\,!}\left(\frac{1.3}{2.4}\nabla_1{}^3+\frac{3}{8}\nabla_1\nabla_2{}^2\right)+\frac{a^5}{5\,!}(\ldots)+\ldots,$$

or

$$\overline{\cos\theta\{e^{\mathbf{a}\triangle}\}}=\tfrac{1}{2}a\left[1+\frac{a^2\nabla^2}{2^2.2}+\frac{a^4\nabla^4}{2^2.4^2.3}+\ldots\right]\nabla_1.$$

Similarly, the reader will find

$$\overline{\sin\theta\{e^{\mathbf{a}\nabla}\}}=\tfrac{1}{2}a\left[1+\frac{a^2\nabla^2}{2^2.2}+\frac{a^4\nabla^4}{2^2.4^2.3}+\ldots\right]\nabla_2.$$

Thus, and recalling that $\nabla_1\mathrm{R}_1+\nabla_2\mathrm{R}_2=\operatorname{div}\mathbf{R}$,

$$\frac{1}{2\pi a}\,\mathrm{S}=\frac{a}{2}\left\{1+\frac{a^2\nabla^2}{2^2.2}+\frac{a^4\nabla^4}{2^2.4^2.3}+\ldots\right\}\operatorname{div}\mathbf{R}.$$

But, if $\mathrm{J}_1(x)=-\mathrm{J}'(x)$ be the Bessel function of order one,

$$\mathrm{J}_1(ix)=\frac{ix}{2}\left(1+\frac{x^2}{2^2.2}+\frac{x^4}{2^2.4^2.3}+\ldots\right).$$

Hence, the required *flux across the circle* receives the compact expression

$$\int_0^{2\pi}\mathbf{aR}\,d\theta=2\pi a^2\left\{\frac{\mathrm{J}_1(ia\nabla)}{ia\nabla}\right\}\operatorname{div}\mathbf{R}.\quad\ldots\ldots\quad(4)$$

It depends only on the *divergence* of the field at the centre, and the bracketed operator contains again only the Laplacean ∇^2 and its iterations.

Since $d\mathbf{a}=a(\mathbf{j}\cos\theta-\mathbf{i}\sin\theta)d\theta$, *the line-integral around the circle* is given by

$$\frac{1}{2\pi a}\int\mathbf{R}\,d\mathbf{a}=\overline{\cos\theta\{e^{\mathbf{a}\nabla}\}}\mathrm{R}_2-\overline{\sin\theta\{e^{\mathbf{a}\nabla}\}}\mathrm{R}_1$$

$$=\frac{a}{2}\left\{1+\frac{a^2\nabla^2}{2^2.2}+\frac{a^4\nabla^4}{2^2.4^2.3}+\ldots.\right\}(\nabla_1\mathrm{R}_2-\nabla_2\mathrm{R}_1),$$

or, in terms of the Bessel function,

$$\int\mathbf{R}\,d\mathbf{a}=r2\pi a^2\left\{\frac{\mathrm{J}_1(ia\nabla)}{ia\nabla}\right\}|\mathrm{curl}\,\mathbf{R}|.\quad\ldots\ldots\quad(5)$$

Thus one and the same operator, applied to the divergence and the curl at the centre, gives the flux across and the circulation around the circle respectively *.

The reason of this coincidence is not hard to see. If $d\sigma$ be an element of area of the circle, the flux and the line-integral are

$$\int\mathbf{aR}\,d\theta=\int\operatorname{div}\mathbf{R}.d\sigma,\quad\int\mathbf{R}\,d\mathbf{a}=\int|\mathrm{curl}\,\mathbf{R}|\,d\sigma,$$

---

* The operand in (5) is the absolute value of curl **R** or **k** curl **R**, if **k** be a unit vector normal to the plane, $\mathbf{k}=\mathrm{V}\mathbf{ij}$.

so that (4) says: the operator $\Phi=2\left\{\frac{J_1(ia\nabla)}{ia\nabla}\right\}$ applied to the scalar div **R** gives the average of this scalar over the area of the circle, and (5) says the same thing about the scalar $|\mathrm{curl}\,\mathbf{R}|$. In fine, we have for the average of any scalar function over *the area* of a circle,

$$\frac{1}{\pi a^2}\int f\,d\sigma=\Phi f=2\left\{\frac{J_1(ia\nabla)}{ia\nabla}\right\}f. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

In much the same way one finds, for the flux of a vector across the surface of a *sphere* of radius a,

$$\int \mathbf{Ra}\,.\,d\sigma=4\pi a^3\left\{\frac{a\nabla\,.\cosh(a\nabla)-\sinh(a\nabla)}{(a\nabla)^3}\right\}\operatorname{div}\mathbf{R}. \quad . \quad . \quad (7)$$

Here a similar remark can be made. For the flux is now equal to the integral of div **R** taken over the *volume* of the sphere. Thus $3\{\}f$ is the volume-average of any scalar f, which is identical with the expression given in the footnote to page 591 and found originally by integration over a of the surface-average. Needless to say, the operator in (7) is again an even function of $a\nabla$ or contains the Laplacean only.

The result (6) can be verified as follows. Since

$$\frac{\partial}{\partial a}f\,d\sigma=a\int_0^{2\pi}f\,d\theta=2\pi a\,\Omega f=2\pi a\Phi f+\pi a^2\frac{\partial\Phi}{\partial a}f,$$

the relation between Φ and the circular operator Ω must be

$$\Phi+\frac{a}{2}\frac{\partial\Phi}{\partial a}=\Omega. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Now, writing for brevity $ia\nabla=x$ and recalling that $J_1=-J'$, we have, by (6), $\partial\Phi/\partial a=-2/a[J''(x)-J'(x)/x]$ and, therefore,

$$\frac{a}{2}\frac{\partial\Phi}{\partial a}+\Phi=-J''(x)-\frac{1}{x}J'(x)=J(x)=\Omega,$$

in agreement with (8).

In order to illustrate the working or handling of the operators Ω, more especially the circular one, we shall now consider a number of more or less simple cases.

Examples.—Let us contemplate the two-dimensional operator. Let, moreover, the field be radially symmetrical, that is to say, $f=f(r)$, where r is the distance from a fixed point O, the centre of the field, while P, any point of the plane, will be the centre of our circle. To avoid confusion, let $r=\overline{\mathrm{OP}}$, while s will stand for the distance of any point of the circle from O. Then, if we take PO as the axis $\theta=0$, the meaning of the formula (II) is

$$\overline{f(s)}=\frac{1}{2\pi}\int_0^{2\pi}f(s)\,d\theta=\Omega f(r)=\{J(ia\nabla)\}f(r), \quad . \quad . \quad . \quad . \quad (9)$$

where
$$s^2=r^2\left[1-\frac{2a}{r}\cos\theta+\frac{a^2}{r^2}\right]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

At the same time,
$$\nabla^2 f=\frac{1}{r}f'+f''=\frac{1}{r}\frac{d}{dr}\left(r\frac{df}{dr}\right). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

To bring into evidence the radius a as parameter of the operator we shall write, whenever necessary, Ω_a.

Ex. 1.—The simplest case, from Ω's point of view, is that of a harmonic function, $\nabla^2 f=0$, which, by (11), and apart from the trivial $f=\text{const.}$, is necessarily $\log r$. Thus, by (9), if O is not inside the circle,

$$\log s=\log r.$$

This property is well known from the theory of the logarithmic potential or function theory. But since it is, at any rate, a direct performance of the operator Ω, we may dwell a moment upon the integral which it actually evaluates. Since, by (10),
$$\log s=\log r+\tfrac{1}{2}\log\left[1-\frac{2a}{r}\cos\theta+\frac{a^2}{r^2}\right],$$

the evaluated integral is
$$\int_0^{2\pi}\log(1+\epsilon^2-2\epsilon\cos\theta)\,d\theta=0, \quad \text{for } \epsilon<1,$$

or, if we put $\kappa=2\epsilon/(1+\epsilon^2)$,
$$\int_0^{\pi/2}\log\left(1-\kappa^2\cos^2\theta\right)d\theta=-r\log\left(1+\epsilon^2\right).$$

This holds for $\epsilon<1$ and as close to unity as we please. Such, then, is the meaning of $\Omega_a \log r=\log r$, for $a<r$.

Now for non-harmonic radial fields. The simplest of these is not $f=r$, but any *even* positive power of r. For while for r the operator $J(ia\nabla)$ leads to an infinite series, the series terminates for any r^{2n}. The results, of course, are correspondingly simple, but for the sake of illustration of Ω a couple of examples of this kind may deserve a moment's attention.

Ex. 2.—Let $f=r^2$. Then, by (11), $\nabla^2 f=6$, $\nabla^4 f=0$, etc. Thus, by (9),
$$\overline{s^2}=\left\{1+\frac{a^2\nabla^2}{6}\right\}f=r^2+a^2,$$

which holds for any a/r, since neither f nor $\nabla^2 f$, etc. have a singularity. This simple result follows, of course, also directly from (10), since $\overline{\cos\theta}=0$. If the reader so prefers, he may consider it just as a verification of theorem (II).

Ex. 3.—Similarly, for $f=r^4$, we have $\nabla^2 f=16r^2$, $\nabla^4 f=64$, and, again by (9),

$$\overline{s^4}=r^4+4a^2r^2+a^4,$$

which may be instantly verified by (10). With equal ease any other even power of r can be treated.

Ex. 4.—Let us now consider the more interesting case $f=r$. This field itself has no singularities, but since $f'=1, f''=0$,

$$\nabla^2 f=\frac{1}{r}, \quad \nabla^4 f=\frac{1}{r^3}, \quad \nabla^6 f=\frac{9}{r^5}, \text{ etc.},$$

so that the Laplacean with all of its iterations, entering into Ω, are infinite at the centre of the field. Thus the circle around $P(r)$ must not contain it, but it may expand so as just to touch it. Substituting the Laplaceans in (9) we have

$$\frac{s}{r}=1+\frac{1}{2^2}\left(\frac{a}{r}\right)^2+\left(\frac{1}{2.4}\right)^2\left(\frac{a}{r}\right)^4+\left(\frac{1.3}{2.4.6}\right)^2\left(\frac{a}{r}\right)^6 +\left(\frac{1.3.5}{2.4.6.8}\right)^2\left(\frac{a}{r}\right)^8+\ldots. \qquad (12)$$

This series converges for all $a<r$, and even for $a=r$, when the circle just touches the centre of the field. Notice that the series differs but slightly from that of the elliptic integral,

$$\frac{2}{\pi}\mathrm{F}(k)=1+\left(\frac{1}{2}\right)^2 k^2+\left(\frac{1.3}{2.4}\right)^2 k^4+\ldots,$$

the factors 3, 3.5, etc. being just shifted forward by one step. This, however, makes all the difference. For, while $\mathrm{F}(1)=\infty$, the series (12) still converges for $r=a$, namely,

$$1+\frac{1}{2^2}+\left(\frac{1}{2.4}\right)^2+\left(\frac{1.3}{2.4.6}\right)^2+\left(\frac{1.3.5}{2.4.6.8}\right)^2+\ldots=\frac{4}{\pi}, \qquad (12.1)$$

so that $\bar{s}=4r/\pi$, for $a=r$, or

$$\Omega_r r=\frac{4}{\pi}.r. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (12'.1)$$

As a verification, we have in this case, by (10), $s=2r\sin\theta/2$, whence $s=4r/\pi$. As a matter of fact, the series (12.1) converges rapidly enough, and (12) for any $a<r$ more rapidly of course *. In passing, the operator

---

* In terms of the elliptic integral $\mathrm{E}(k)$, of the second kind, (12) can be written

$$\bar{s}=\Omega_a r=\frac{2}{\pi}(a+r)\mathrm{E}(k), \quad k=2\sqrt{\frac{a}{r}}\Big/\left(1+\frac{a}{r}\right).$$

Ω has given us a series for $4/\pi$ (a quaint distortion of the F-series), which to the best of my knowledge is new. The result (12′.1) reads : $\Omega_r = 4/\pi$ with r as operand. Since the field $f=r$ is not harmonic, Ω cannot, of course, be an idemfactor for it, but it just multiplies it (in the case $a=r$) by the mere number $4/\pi$.

In much the same way any odd power of r will give rise to an infinite series and does not present any essential difference, the singularity being only thrown from $\nabla^2 f$ itself to some iteration of the Laplacean, *e. g.*, for $f=r^3$, $\nabla^2 f = 9r$, so that $\nabla^4 f$, etc. are infinite at the field's centre. Examples of this kind, as also those for negative powers of r, may be left to the care of the reader.

Ex. 5.—Consider $f=e^{r/\lambda}$, where λ is a positive or negative constant. Here,

$$\nabla^2 f = \frac{1}{\lambda^2}\Big(1+\frac{\lambda}{r}\Big)f, \quad \nabla^4 f = \frac{1}{\lambda^4}\Big(1+\frac{2\lambda}{r}\Big)f, \text{ etc.,}$$

and (9) gives

$$\bar{f} = \frac{1}{2\pi}\int_0^{2\pi} e^{s/\lambda}\, d\theta$$

$$= \Big[\mathrm{J}\Big(i\frac{a}{\lambda} + \frac{a^2}{4\lambda r}\Big\{1+\frac{2}{4^2}\Big(\frac{a}{\lambda}\Big)^2 + \frac{3}{4^2.6^2}\Big(\frac{a}{\lambda}\Big)^4 + \dots\Big\}\Big]e^{r/\lambda}. \quad . \quad (13)$$

In particular, for $\lambda = \pm a$,

$$\overline{e^{\pm s/a}} = \Big[\mathrm{J}(i) + \frac{a}{4r}\Big(1+\frac{2}{4^2}+\frac{3}{4^2.6^2}+\dots\Big)\Big]e^{\pm r/a}, \quad . \quad . \quad (13.1)$$

respectively, where $\mathrm{J}(i) = 1{\cdot}2661$ *, a number *not* simply related to the ubiquitous π, though but slightly smaller than $4/\pi = 1{\cdot}2732$.

The purely numerical series-factor of $a/4r$, again converges rapidly to another such number. If the centre O of the field is on the circle, nothing extraordinary happens. We then have, simply, (13.1) with $r/a = \pm 1$, which thus evaluates the integrals

$$\int_0^{\pi} e^{\pm 2\sin\alpha}\, d\alpha = \Big[\mathrm{J}(i) \pm \frac{1}{4}\Big(1+\frac{1}{8}+\frac{1}{192}+\dots\Big)\Big]\pi e^{\pm 1}, \quad . \quad (13.2)$$

whence, *e. g.*,

$$\frac{1}{\pi}\int_0^{\pi}\cosh(2\sin\alpha)\, d\alpha = \mathrm{J}(2i) = \mathrm{J}(i)\cosh 1 + \frac{1}{4}\Big(1+\frac{1}{8}+\frac{1}{192}+\dots\Big)\sinh 1,$$

which, in fact, approaches rapidly the tabulated value $\mathrm{J}(2i) = 2{\cdot}2796$.

---

* *Cf.* 'Jahnke and Emde's Tables,' p. 130.

Ex. 6.—A noteworthy case arises when the radial field is a Bessel function of r/λ or, with the constant λ as unit distance, simply

$$f(r)=\mathrm{J}(r).$$

Then $f''+\frac{1}{r}f'+f=0$ (Bessel's equation), so that $\nabla^2 f=-f$ or $\nabla=i$, and our operator becomes a simple numerical multiplicator (factor), $\Omega=\mathrm{J}(ia\nabla)=\mathrm{J}(a)$, and we have the elegant relation

$$\overline{\mathrm{J}(s)}=\mathrm{J}(a)\,.\,\mathrm{J}(r), \qquad (14)$$

where r is again the distance of P from the centre of the field O, and s the distance of O from any point of the circle. In this case neither f nor $\nabla^2 f=-f$, $\nabla^4 f=f$, etc. have a singularity at O, assuming there simply the values ± 1. Thus (14) holds also when the circle contains the field's centre. In particular, if O is on the circle, $a=r$, and we have

$$\overline{\mathrm{J}(s)}=\mathrm{J}^2(r),$$

which evaluates the rather complicated integral

$$\frac{1}{\pi}\int^{\pi}\mathrm{J}(2r\sin\alpha)\,d\alpha=\mathrm{J}^2(r). \qquad (14.1)$$

A radial Bessel field is thus the next simple thing to a harmonic field. For the latter Ω is an idemfactor, and for the former an idemfactor multiplied by a number, the Bessel function $\mathrm{J}(a)$ of the chosen radius.

These examples of planar fields may suffice. Three-dimensional fields can be treated in much the same way. If we consider radially symmetrical fields, $f=f(r)$, where r is the distance of P (and s that of any point on the sphere a about P) from a fixed centre O, then (III) assumes the form

$$\overline{f(s)}=\left\{\frac{\sinh(a\nabla)}{a\nabla}\right\}f(r), \qquad (15)$$

where

$$\nabla^2 f=\frac{2}{r}f'+f'' \qquad (16)$$

and s is as in (10). For a harmonic field $1/r$ (with O not within the sphere) $\overline{f(s)}=f(r)$, of course [*]. If $f(r)$ is any even power of r, the operator in (15) consists, as before, of a finite number of terms, without singularities, *e. g.*, for $f=r^2$, $\nabla^2 f=6$, $\nabla^4 f=0$, etc., so that

$$\overline{s^2}=a^2+r^2.$$

---

[*] If O is inside the sphere, then s^{-1}=const., *i. e.*, independent of the position of O inside the sphere, the well-known property of the Newtonian potential of a homogeneous spherical sheet.

Similarly, for $f=r^4$, $\nabla^2 f=20r^2$, $\nabla^4 f=120$, $\nabla^6 f=0$, etc., and

$$s^4=a^4+\frac{10}{3}a^2r^2+r^4,$$

and so on, for any position of the centre O. The series terminates also for any odd (positive) power of r, but is not free from singularities. It may suffice to consider the first power.

Ex. 7.—Radial field, $f=r$. Since $f'=1, f''=0$, we have $\nabla^2 f=2/r$, $\nabla^4 f=0$, so that the series terminates again soon enough and we have, by (15),

$$\bar{s}=r\left(1+\frac{1}{3}\frac{a^2}{r^2}\right).$$

This holds only for $a<r$, since $\nabla^2 r$ has a singularity at O. But it remains valid for $a=r$ (O on sphere), giving

$$\bar{s}=\frac{4}{3}r.$$

This may be instantly verified by (10), which reduces to $s=2r\sin\theta/2$, giving

$$\frac{\bar{s}}{r}=\frac{2\pi}{4\pi}\int_0^\pi 2\sin\frac{\theta}{2}\sin\theta\,.\,d\theta=\frac{4}{3}.$$

Other odd powers, as *e. g.*, r^3, giving $\nabla^2 f=12r$, $\nabla^4 f=24/r$, etc., and fractional or negative powers may be left to the care of the reader.

Ex. 8.—Analogously to Ex. 6, consider the radial field

$$f(r)=\frac{\sin hr}{r}.$$

Then, by (16),

$$\nabla^2 f=f \quad \text{or} \quad \nabla^2=1,$$

so that the operator becomes a simple multiplier, $\Omega=\sinh a/a$, and we have the remarkable result

$$\sinh s/s=\frac{\sinh a}{a}\cdot\frac{\sinh r}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Since neither f nor any of its Laplaceans has a singularity at O, this formula holds for any position of O, outside, on, or inside the sphere. If O is on the sphere, we have $\overline{\sinh s/s}=\sinh^2 r/r^2$, and if it coincides with the centre of the sphere ($r=0$), (17) becomes, of course, an identity for any a. In (17) s has the meaning (10), so that this formula evaluates an integral over a sphere for any value of the ratio of a and r.

More generally, for any field $f=f(x, y, z)$ satisfying the equation

$$\nabla^2 f=m^2 f, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

where m is a constant, the operator Ω becomes a multiplier, and

$$\bar{f}=\frac{\sinh ma}{ma}\cdot f, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (18.1)$$

where f is the field at the centre of the sphere a. If m^2 is positive, the factor of f steadily increases with increasing radius. And if $m^2<0$, say $m^2=-\kappa^2$, the factor is an oscillatory function of the radius, $\sin \kappa a/\kappa a$, so that the mean $\bar{f}$ vanishes periodically for all $a=n\pi/\kappa$.

These few examples ought to suffice for illustrating the concepts of the circular integrator $\mathrm{J}(ia\nabla)$ and the spherical integrator $\dfrac{\sinh a\nabla}{a\nabla}$ and for stimulating, perhaps, their application.

129 Seneca Parkway,
Rochester, N.Y.

LX. *Note on Paper on " Calculation of Triode Constants "* (Phil. Mag., June 1939, p. 709).

To the Editors of the Philosophical Magazine.

GENTLEMEN,—

THE paper entitled " Calculation of Triode Constants " by Dr. J. H. Fremlin (Phil. Mag., June 1939, p. 709) is a most valuable contribution to the subject. There is, however, an unstated assumption in that the position of the equivalent diode plane, given by equation (XIII.) is assumed to be constant at any grid bias, whereas it has necessarily this value only at the grid voltage

$$\left(\frac{lg}{la}\right)^{4/3} V_A.$$

Sometime ago the writer made a similar calculation on the following lines :—Suppose that a very high μ grid at a potential V_L is located between grid and anode. Then the current flow and potential distributions both in the cathode grid space and in the grid to anode space can be calculated, assuming that the electrons are projected into the grid to anode space with an initial velocity given by V_L. The limiting values of the intensity on each side of the grid will determine Q_g, the charge per cm.² induced on the grid. Now suppose the high μ grid to be replaced by a low μ grid, then the actual potential of the grid wires, V_g, will be depressed below the potential V_L by an amount

$$-\frac{4\pi Q_g}{K},$$

where

$$K=\frac{2\pi N}{\log_e \frac{1}{\pi N\delta}},$$

N=No. of turns per cm.,

δ=Diam. of grid wires in cm.,

K is a constant independent of the grid location.

The above relation can be shown to be true for the electrostatic case, and is assumed to be true on using the value of Q_g in the case of current flow. We then have I_A/cm.² a known function of V_L, and V_L of V_g.

Suppose now that Q_g is a linear function of V_g, and that Q_g equals zero when V_g equals some voltage V_f. The following relationship follows :—

$$I_A/\text{cm.}^2 = \frac{2{\cdot}34\times10^{-6}\left(\frac{V_A}{\mu}+V_g\right)^{3/2}}{X_1^2\left[1+\frac{1}{\mu}\frac{V_A}{V_f}\right]^{3/2}} \text{ amperes/cm.}^2. \quad . \quad . \quad (1)$$

As an example, if the effect of space charge in the grid to anode space is neglected

$$V_f = \frac{\frac{3}{4}X_1}{\frac{3}{4}X_1+X_2},$$

where X_1=grid to cathode gap, X_2=grid to anode gap (in cm.), so that

$$I_A/\text{cm.}^2 = \frac{2{\cdot}34\times10^{-6}\left(\frac{V_A}{\mu}+V_g\right)^{3/2}}{X_1^2\left[1+\frac{1}{\mu}\frac{X_1+\frac{4}{3}X_2}{X_1}\right]^{3/2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In actual fact, however, the relation between Q_g and V_g is not quite linear.

Let Q_2=charge induced on the anode side of the grid, then it can be shown that if V_L is small compared with V_A

$$Q_2 = a + bV_L + cV_L^{3/2},$$

where the third term arises as an effect of space charge between grid and anode and is proportional to anode current. Clearly when V_L=zero, as near cut-off, the conditions giving equation (2) are satisfied, but when $V_g = \left(\frac{X_1}{X_1+X_2}\right)^{4/3} V_A$, we must satisfy Dr. Fremlin's equation.

As a final equation, I obtain

$$I_A/\text{cm.}^2 = \frac{2{\cdot}34\times10^{-6}\left[\frac{A}{\mu}+V_g\right]^{3/2}}{X_1^2\left\{\left[1+\frac{X_1+\frac{4}{3}X_2}{\mu X_1}\right]+\frac{1-\left(\frac{X_1+\frac{4}{3}X_2}{X_1}\right)\left(\frac{X_1}{X_1+X_2}\right)^{4/3}}{\mu\left[\frac{X_1}{X_1+X_2}\right]^2\left[1+\frac{1}{\mu}\left(\frac{X_1+X_2}{X_1}\right)^{4/3}\right]^{1/2}}\left(\frac{1}{\mu}+\frac{V_g}{V_A}\right)^{1/2}\right\}^{3/2}},$$

showing that the equivalent diode plane varies with V_g.

It should be noted that the deviations from Dr. Fremlin's equation or from equation (2) are more apparent at large values of the ratio $\frac{X_2}{X_1}$.

Yours faithfully,
S. Rodda.

Valve Laboratory,
Cosmos Manufacturing Co. Ltd.

LXI. *Erratum to the Paper "On the Interrelation of Molecular Constants for Diatomic Molecules.—II."* (Phil. Mag., March 1940, p. 298).

To the Editors of the Philosophical Magazine.

Gentlemen,—

In a short note at the end of § 1 (p. 299), it was incorrectly stated that the bond force constant k in Sutherland's relation $D = kr_e^2/mn$ was that used in one of my earlier papers. Sutherland has used the accepted definition of k and his result does not require modification.

Yours faithfully,
R. A. Newing.

April 2, 1940.

LXII. *Notices respecting New Books.*

Theoretical Mechanics. By C. J. Coe. [Pp. xiii+555.] (The MacMillan Company, New York. Price 5$.)

In this book the author employs the methods of vector analysis which have come into more general use in the study of mechanics in recent years. The vector analysis appears almost incidentally in the development of the mechanics: the rectilinear motion of a particle involves only scalar quantities; a brief exposition of Euclidean and analytic geometry employs vector algebra; the curvilinear motion of a particle and the motion of a rigid body use the vector function of a scalar; the statics of particles and rigid bodies has as its basis a discussion of systems of sliding vectors; the kinetics of particles and of a rigid body involve the application of the linear vector function; and a discussion of vector calculus leads to an account of potential theory. Numerous exercises appear in the book which will be very helpful to first and second year students at the universities.

The Mathematical Papers of Sir Wm. Rowan Hamilton.—Vol. II. *Dynamics.* Edited by Professors Conway and McConnell. [Pp. xv+656.] (Cambridge University Press, 1940. Price 70*s.* net.)

The Royal Irish Academy is indeed raising to the memory of one of the greatest of the Irish school of mathematicians *monumentum aere perennius.* They have already sponsored a volume dealing with Hamilton's contribution to Geometrical Optics, and the noble and massive quarto under review is devoted to his work on Dynamics, which, as published here, falls into three main groups—the group which centres on his two great papers of 1834 and 1835 on a general method in dynamics, in which the main properties of his characteristic functions are discussed. This section contains several papers hitherto unpublished and, in the shape of a British Association Report, a brief discussion of the application to dynamics of a general mathematical method previously applied to optics.

The second section, which is composed chiefly of material hitherto existing in manuscript form, is concerned with a new calculus, the Calculus of Principal Relations, which is "an extension of the method of his principal function to obtain integrals of total and even partial differential equations." The third section consists of seven papers on the dynamics of light. Here again, much of this section is composed of manuscript hitherto unpublished. Indeed, the major portion of this volume consists of such unpublished matter, no fewer than about 450 of the 650 pages of the volume being devoted to papers of which the sources are given as "Manuscript." This material, though much of it is tentative and incomplete, is specially valuable as showing Hamilton's method of feeling his way by working out laborious examples, testing every step in a generalization by a particular application. These papers are specially valuable as showing Hamilton's method of attack, and the editors are to be congratulated on their decision to print them.

The papers on the dynamics of light are, historically, of great interest, and, as the editors remark, Hamilton's insistence on the classical methods of a particle-ether is in sharp contrast to that of his contemporary MacCullagh. Here Hamilton anticipated some of Cauchy's results and came to the conclusion that, in the existing state of experimental physics, the precision of the experiments made was not sufficiently great to make it possible to decide on the question of the direction of the light vector.

The volume is magnificently produced and the printing, which has had to face a symbolism demanding the most meticulous attention from the proof-correctors, touches the highest of the high standards set by the Cambridge University Press. The complete edition will form a lasting and worthy memorial of one of the greatest, most original, and most versatile minds of his generation.

Principles of Electricity and Magnetism. By Prof. HARNWELL. (London: McGraw-Hill Publishing Co., Ltd., 1938. Price 30*s.*)

THE author has provided a course which is well designed to suit the needs of a physicist or engineer working in the graduate school. Emphasis is placed on the experimental rather than on the theoretical, on the classical rather than on the "modern" side, though discussions of discharge phenomena and of electronics in general are included in the book.

An elementary knowledge of electricity and of the calculus is pre-supposed, otherwise the book is self-contained, a mathematical appendix being devoted to such topics as Fourier analysis, differential equations, and that amount of vector theory necessary for the comprehension of a course of experimental electricity and magnetism of honours standard.

The scope of the book may be estimated from the titles of its principal chapters, which deal with Electrostatic Energy and Dielectrics, the Physical Characteristics of Dielectrics and of Conductors, Direct and Alternating Current Circuits, Chemical, Thermal and Photo-electric Effects, Thermionics, Conduction in Gases, Electromagnetic Effects, Magnetism, Inductive Circuits, Vacuum-tube circuits and Radiation.

The author has no special hobbies to expound, handles his topics thoroughly and clearly, and has the teacher's flair for appreciating and elucidating those topics about which the student is apt to find himself in difficulties. A. F.

The Concepts of the Calculus. By C. B. BOYER. [Pp. vi + 346.] (Columbia University Press. England: Oxford University Press, 1939. Price 18*s.* 6*d.*)

THERE is no lack of literature concerning the origin and development of the differential and integral calculus; the bibliography which is appended to this treatise runs to nearly five hundred items. Nevertheless, until the publication of this work, there has not appeared a critical and evolutionary story of the calculus dealing with its beginnings in the mists of antiquity, the contributions of the Middle Ages leading to a "century of anticipation," the work of Newton and Leibniz, and the two periods designated by the author as the period of indecision and that of the rigorous formulation.

The author has dealt with difficult logical matters clearly and authoritatively; indeed, it would not be easy to improve the picture which he gives of the gradual strengthening of the logical bases of the calculus. Not that some of our greatest minds worried much about the details of the fundamentals. It is told of Kelvin that he introduced a course on the higher mathematics by writing dx/dt on the blackboard and asking for its meaning. He grew more and more impatient as he received variants of the statement that it was the limiting value, as t

approached zero, of $\delta x/\delta t$. "That's what Todhunter would say," he remarked; "Does nobody know that dx/dt means velocity?"

As the author comments, Lord Kelvin considered mathematics to be the etherealization of common sense.

The book may be heartily commended to all readers interested in the history of mathematics.

Modern Science; a Study of Physical Science in the World Today. By Prof. H. Levy. [Pp. x+736.] (London: Hamish Hamilton, 1939. Price 21*s.*)

Recent years have seen a flood of books pouring from the world's presses which endeavour to tell the citizen something of the aims, trends, and results of modern science. Professor Levy's book is one of the latest of a very long list, and is one of the best of its kind.

It differs, indeed, from many of its predecessors in the emphasis continually laid on the relation of scientific advances to social needs; in its insistence on the reasons for a particular outburst of scientific discovery at a particular point in the development of man's history; in its correlation of scientific advance and political and economic history.

But the book is much more than this; it does not merely touch, in passing, on the details of the story of scientific discovery—it provides an account of those discoveries which may be read with pleasure by a layman, and which is expounded with all the skill and clarity which we are accustomed to expect from Professor Levy. He has, for example, given his readers an elementary account of relativity and of recent developments in physical science which could hardly be bettered; and his chapter on Illumination is a model of clear exposition.

Some will dissent from Professor Levy's interpretation of history. That is a secondary matter. The book will have amply fulfilled its purpose if it stimulates the citizen to embark on that hardest of all hard tasks—thinking for one's self.

The History of the Institution of Electrical Engineers, 1871–1931. By Rollo Appleyard. [Pp. 342.] (Published by the Institution of Electrical Engineers, London, 1939. Price 18*s.* 6*d.*)

There is a vast mass of important scientific history hidden away in the archives of our great Institutions, and he who searches therein is likely to chance upon treasure of many kinds. But the searcher must possess many and varied qualifications. He must not be afraid of the dry details of search and research. Nothing is more annoying than to find a history rich in flowery talk and erudite generalization, and forever letting one down by its lack of those facts and dates, dry in themselves, but very necessary to those who would set the story, whatever it may be, against the background of its times. On the other hand, a mere catalogue of facts and dates forms the dullest of all compilations, though its usefulness may considerably exceed that of its more pretentious brother.

Mr. Appleyard is admirably qualified, by training and habit, for the task which he has completed with such conspicuous success. He is not afraid of detail and of lists. The reader who would seek the title of the Kelvin Lecture for 1920, the name of the Salomons scholar for 1898, a list of the first wiring rules of the Society of Telegraph Engineers, the date of Lord Kelvin's famous pronouncement on the relative merits of direct and alternating current supply, will find his questions clearly and accurately answered. But Mr. Appleyard does much more than this. He has told the fascinating tale of the development of applied electrical science, which has so much influenced, and in its turn been influenced by, the development of the Institution. He has covered in detail the period from the early efforts at telegraphy to our own days of wireless and of television ; and he has told the story with a wealth of historical knowledge and personal detail which makes his book one of absorbing interest.

The figures of the great pioneers are sketched vigorously and firmly, and there are scores of incidental sketches which help the reader to visualize a past, now historic, but still within the memory of living man. The book is a notable contribution to scientific history, and one which every electrical engineer may read with profit.

A. F.

The Kelley Statistical Tables. By T. L. KELLEY. [Pp. 136.] (The Macmillan Company, New York. Price 20*s*. net.)

THIS collection of tables has grown out of the demands of a particular problem for which existing tables were inadequate in that their use involved a high order of interpolation in order to secure the desired accuracy. The principal table is a normal distribution, simple correlation, and probability function table, in which values of x, the distance from the mean, and of z, the ordinate at the point of dichotomy, are listed to eight decimal places at intervals of 0·0001 in p, the larger proportion in a dichotomized unit normal distribution. The Lagrangian interpolation coefficients which may be employed in using the chief table appear in subsidiary tables and are given to ten or more decimal places. Other useful information given includes notes on the use of the tables, four place χ^2 functions, and eight place square roots. The arrangement and the printing of the tables are excellent, and the gratitude of many statistical workers will be due to the writer for the labour involved in their compilation.

Advanced Calculus. By W. B. FITE. [Pp. xii+399.] (The Macmillan Company, New York. Price 21*s*. net.)

THE book supplies an introduction to mathematical analysis which will be of value to students proposing to specialize in mathematics. Following the opening chapter, which is devoted to a consideration of systems of real numbers, the subject is developed in the customary manner and discussions of the usual topics and theorems are given. The concluding chapters provide short introductions to the calculus of variations and to functions of a complex variable.

Complex Variable and Operational Calculus. By N. W. McLACHLAN. [Pp. xi +355.] (Cambridge University Press. Price 25*s*. net.)

THE difficulties involved in the original and essentially experimental operational procedure due to Heaviside have been overcome during the last twenty-five years, and the author presents a modern treatment of the operational method based upon a particular case of the Mellin inversion theorem. Bearing in mind that the mathematical technologist, to whom the book is addressed, is not primarily concerned with the rigour of the methods he employs, the writer has attempted to strike a happy mean between the demands of the pure mathematician and of the practical worker.

For the benefit of readers not acquainted with complex variable theory, an exposition of the relevant portions of this subject is first given, and this is followed by the development of the operational calculus and procedure. The method is exemplified in a host of technical applications of which a few may be cited : electrical circuits, radio and television receivers, cable telegraphy, wave filters, condenser microphone, loud speaker horn, and heat diffusion. In a number of appendices are found discussions of certain points arising in the text and an extensive bibliography.

The method of presentation is stimulating, and the volume will be studied by all interested in technical problems and their solution.

Introduction to Bessel Functions. By F. BOWMAN. [Pp. x+135.] (Longmans, Green & Co. Ltd. Price 10*s.* 6*d.* net.)

THIS attractive volume will be welcomed by students of mathematics and of physics and engineering. For the mathematical student it will serve as a useful introduction, and will be read prior to the study of standard treatises on Bessel functions ; while the non-mathematical specialist, who may be awed by the standard works, will greatly appreciate the simple treatment offered here.

Five chapters of the book are devoted to the study of zero order functions, and functions of any order are treated in the sixth chapter. The importance of Bessel functions is illustrated in various applications which include the problems of the vibrating circular membrane, the hanging chain, the flow of alternating current in a wire of circular section, and the critical length of a vertical rod. Many examples for solution appear in the text, and the book is reasonably priced.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

INDEX TO VOL. XXIX.

ALLAN (G. E.), the musical scale of the highland bagpipe, 154.

Band (W.) on Flint's five-dimensional theory of the electron, 548.

Barker (H.) and Wheatcroft (E. L. E.) the transition from glow to arc discharge, 1.

Becker (J. A.) and Moore (G. E.), thermionic emission, migration, and evaporation of barium on tungsten, 129.

Books, new:—A. S. Eve, Rutherford: being the Life and Letters of the Rt. Hon. Lord Rutherford, O.M., 103; H. Semat, Introduction to Atomic Physics, 103; S. Chapman and T. G. Gowling, The Mathematical Theory of Non-Uniform Gases, 104; J. H. Hibben, The Raman Effect and its Chemical Applications, 208; H. C. Plummer, Probability and Frequency, 310; A. W. Barton, A Text Book on Light, 311; J. S. Rogers, Physics for Medical Students, 311; M. H. A. Newman, Elements of the Topology of Plane Sets of Points, 312; Terrestrial Magnetism and Electricity, edited by J. A. Fleming, 415; A. I. Vogel, A Text-Book of Quantitative Inorganic Analysis Theory and Practice, 416; Erwin Meyer, Electro-acoustics, 517; J. Reilly and W. N. Rae, Physico-chemical Methods, 517; A. S. Eddington, The Philosophy of Physical Science, 518; D. A. MacInnes, The Principles of Electrochemistry, 519; N. A. McKenna, Theoretical Electrochemistry, 520; C. J. Coe, Theoretical Mechanics, 604; The Mathematical Papers of Sir Wm. Rowan Hamilton.—Vol. II. Dynamics. Edited by Conway and McConnell, 604; Harnwell, Principles of Electricity and Magnetism, 605; C. B. Boyer, The Concepts of the Calculus, 605; H. Levy, Modern Science; a Study of Physical Science in the World Today, 606; R. Appleyard, The History of the Institution of Electrical Engineers, 1871–1931, 606; T. L. Kelley, The Kelley Statistical Tables, 607; W. B. Fite, Advanced Calculus, 607; N. W. McLachlan, Complex Variable and Operational Calculus, 607; F. Bowman, Introduction to Bessel Functions, 608.

Bowden (S. T.) and Morgan (A. R.), the temperature variation of viscosity in non-polar and di-polar inorganic liquids, 367.

Brønsted (J. N.), the fundamental principles of energetics, 449.

Brown (A. E.), the effect of a magnetic force on high frequency discharges in pure gases, 302.

Carey-Foster method of measuring a mutual inductance in terms of a capacity, note on the, 162.

Cockett (A. H.) and Ferguson (A.), the specific heat of water and of heavy water, 185.

Crowther (J. A.), Liebmann (H.), and Jones (R.), experiments on the action of ionizing radiations on the electrophoretic mobility of colloid particles, 391.

2 U

Dielectric constant and electrical conductivity of gases and vapours ionized by X-rays at ultra-high radio frequency, 353.

—— —— of ionized air.—III., 344.

Differential operators acting as integrators, 586.

Diffraction of heat, on some problems in the, 93.

Distributions of temperature and vapour pressure around a horizontal wet cylinder, note on the, 274.

Dube (G. P.) and Levine (S.) on the mutual electrical energy of two colloidal particles ; general theory, using approximate Debye-Hückel equation, 105.

Effect of a magnetic force on high frequency discharges in pure gases, 302.

Effusion phenomena in relativistic quantum statistics, 85.

Electric field produced by a point-charge located outside a dielectric wedge, 36.

Electron, on Flint's five-dimensional theory of the, 548.

Equations of motion of point electrons deduced from a variational principle, 495.

Error in paper on " Certain vibration problems solved by means of an analogous problem in heat conduction," Phil. Mag. xxii. p. 1079, 100.

Experiments on the action of ionizing radiations on the electrophoretic mobility of colloid particles, 391.

Ferguson (A.) and Cockett (A. H.), the specific heat of water and of heavy water, 185.

Flint (H. T.), the theory of the electric charge and the quantum theory.—Part I., 330 ; Part II., 417.

Flint's five-dimensional theory of the electron, on, 548.

Fundamental principles of energetics, 449.

Gogate (D. V.), effusion phenomena in relativistic quantum statistics, 85.

Gravity and the rotation of the earth, 471.

Green (G.), error in paper on " Certain vibration problems, solved by means of an analogous problem in heat conduction," Phil. Mag. xxii. p. 1097, 100.

Guest (J. J.) on the proposition that " shearing stress alone is effective in producing elastic failure," 289.

Haidinger's rings in non-uniform plates, 508.

Harmonic analysis, the method of symmetrical components applied to, 66.

Harper (W. R.), a rigorous theory of the recombination of both small and large ions in gases at high pressures, 434.

Hartmann (J.) and Lazarus (F.), the radiation of the acoustic air-jet generator derived from direct observation of the amplitude of the aerial vibrations in the oscillator, 140.

Haseltine (W. R.), Varney (R. N.), and Loeb (L. B.), the rôle of ionization by positive ions in spark breakdown, 379.

Havelock (T. H.), waves produced by the rolling of a ship, 407.

Hoare (F. E.), the radiation correction in calorimetry, 52.

Hollingdale (S. H.), stability and configuration of the wakes produced by solid bodies moving through fluids, 209.

Huxley (L. G. H.), the propagation of electromagnetic waves in an atmosphere containing free electrons, 313.

Hystero-differential equation arising in a probability problem, 75.

Interaction between radiation and matter, 169.

Interrelation of molecular constants for diatomic molecules.—II., 298.

Iron-cored coils of high efficiency, 485.

Iso-dielectric lines and geological structure, 32.

Jaeger (J. C.), magnetic screening by hollow circular cylinders, 18.

Jones (R.), Crowther (J. A.), and Liebmann (H.), experiments on the action of ionizing radiations on the electrophoretic mobility of colloid particles, 391.

Kar (K. C.), interaction between radiation and matter, 169.

—— on the proton-proton scattering, 200.

Khastgir (S. R.) and Rahman (Shah Md. F.), dielectric constant of ionized air.—III., 344.

—— ——, the dielectric constant and electrical conductivity of gases and vapours ionized by X-rays at ultra-high radio frequency, 353.

Krishnan (R. S.), note on "The theory of light-scattering," 515.

Lamont (H. R. L.), theory of resonance in microwave transmission lines with discontinuous dielectric, 521.

Lazarus (F.) and Hartmann (J.), the radiation of the acoustic air-jet generator derived from direct observation of the amplitude of the aerial vibrations in the oscillator, 140.

Lenihan (J. M. A.) and Richardson (E. G.), observations on edge tones, 400.

Levine (S.) and Dube (G. P.) on the mutual electrical energy of two colloidal particles: general theory, using approximate Debye-Hückel equation, 105.

Lewis, (T.) the equations of motion of point electrons deduced from a variational principle, 495.

Liebmann (H.), Crowther (J. A.), and Jones (R.), experiments on the action of ionizing radiations on the electrophoretic mobility of colloid particles, 391.

Light-scattering, on the theory of, 148.

Loeb (L. B.), Varney (R. N.), and Haseltine (W. R.), the rôle of ionization by positive ions in spark breakdown, 379.

Lowan (A. N.) on some problems in the diffraction of heat, 93.

Löwy (H.) on iso-dielectric lines and geological structure, 32.

Magnetic screening by hollow circular cylinders, 18.

Marro (M.), secrecy by inversion of syllables, 205.

Mass and force in Newtonian mechanics—Addendum to "Mass I." and "Mass II.," 477.

Mechanical analogy to the motion of electrons in gases, 47.

Method of symmetrical components applied to harmonic analysis, 66.

Moore (G. E.) and Becker (J. A.), thermionic emission, migration, and evaporation of barium on tungsten, 129.

Morgan (A. R.) and Bowden (S. T.), the temperature variation of viscosity in non-polar and dipolar inorganic liquids, 367.

Musical scale of the highland bagpipe, 154.

Mutual electrical energy of two colloidal particles: general theory, using approximate Debye-Hückel equation, 105.

Newing (R. A.) on the interrelation of molecular constants for diatomic molecules.—II., 298; erratum, 603.

Note on a method for the comprehensive analysis of complex power plant, 568.

—— —— paper on "Calculation of triode constants" (Phil. Mag., June 1939, p. 709), 601.

—— —— the Carey-Foster method of measuring a mutual inductance in terms of a capacity, 162.

—— —— the distributions of temperature and vapour pressure around horizontal wet cylinder, 274.

—— —— the Schottky effect, 16.

Note on "The theory of light-scattering," 515.

Observations on edge tones, 400.

Ohlin (P.), a "scale of three" counter, 285.

On some problems in the diffraction of heat, 93.

Owen (E. A.), the structure of meteoritic iron, 553.

Parthasarathy (S.) on the theory of light-scattering, 148.

Partridge (G. F.), iron-cored coils of high efficiency, 485.

Pendse (C. G.), gravity and the rotation of the earth, 471.

—— on mass and force in Newtonian mechanics—Addendum to "Mass I." and "Mass II.," 477.

Pipes (L. A.), the method of symmetrical components applied to harmonic analysis, 66.

Powell (R. W.), note on the distributions of temperature and vapour pressure around a horizontal wet cylinder, 274.

Principles of a mechanical method for calculating regression equations and multiple correlation coefficients and for the solution of simultaneous linear equations, 258.

Probability problem, on a hystero-differential equation arising in a, 75.

Propagation of electromagnetic waves in an atmosphere containing free electrons, 313.

Proton-proton scattering, on the, 200.

Radiation correctionin calorimetry, 52.

—— of the acoustic air-jet generator derived from direct observation of the amplitude of the aerial vibrations in the oscillator, 140.

Rahman (Shah Md. F.) and Khastgir (S. R.), dielectric constant of ionized air.—III., 344.

—— ——, the dielectric constant and electrical conductivity of gases and vapours ionized by X-rays at ultra-high radio frequency, 353.

Rajagopalan (V. S.) and Raman (C. V.), Haidinger's rings in non-uniform plates, 508.

Raman (C. V.) and Rajagopalan (V. S.), Haidinger's rings in non-uniform plates, 508.

Relativistic quantum statistics, effusion phenomena in, 85.

Resonance in microwave transmission lines with discontinuous dielectric, theory of, 521.

Rice (S. O.), the electric field produced by a point-charge located outside a dielectric wedge, 36.

Richardson (E. G.) and Lenihan (J. M. A.), observations on edge tones, 400.

Rigorous theory of the recombination of both small and large ions in gases at high pressures, 434.

Rodda (S.), note on paper on "Calculation of triode constants" (Phil. Mag., June 1939, p. 709), 601.

Rôle of ionization by positive ions in spark breakdown, 379.

"Scale of three" counter, 285.

Schottky effect, a note on the, 16.

Schumann (T. E. W.), the principles of a mechanical method for calculating regression equations and multiple correlation coefficients and for the solution of simultaneous linear equations, 258.

Secrecy by inversion of syllables, 205.

"Shearing stress alone is effective in producing elastic failure," on the proposition that, 289.

Silberstein (L.), differential operators acting as integrators, 586.

—— on a hystero-differential equation arising in a probability problem, 75.

Some impedance characteristics of tapped resonant-line circuits, 572.

Specific heat of water and of heavy water, 185.

Stability and configuration of the wakes produced by solid bodies moving through fluids, 209.

Stresses in a circular ring supported on frictionless radial pins, 541.
Structure of meteoritic iron, 553.
Sutton Pippard (A. J.), the stresses in a circular ring supported on frictionless radial pins, 541.
Temperature variation of viscosity in non-polar and dipolar inorganic liquids, 367.
Theory of resonance in microwave transmission lines with discontinuous dielectric, 521.
—— —— the electric charge and the quantum theory.—Part I., 330; Part II., 417.
Thermionic emission, migration, and evaporation of barium on tungsten, 129.
Transition from glow to arc discharge, 1.
Triode constants," note on paper on "Calculation of, 601.
Tyler (E.), a note on the Carey-Foster method of measuring a mutual inductance in terms of a capacity, 162.
Varney (R. N.), Loeb (L. B.), and Haseltine (W. R.), the rôle of ionization by positive ions in spark breakdown, 379.
Walshaw (A. C.), note on a method for the comprehensive analysis of complex power plant, 568.
Waves produced by the rolling of a ship, 407.
Wheatcroft (E. L. E.), a note on the Schottky effect, 16.
—— and Barker (H.), the transition from glow to arc discharge, 1.
Willshaw (W. E.), some impedance characteristics of tapped resonant-line circuits, 572.
Yarnold (G. D.), a mechanical analogy to the motion of electrons in gases, 47.

END OF THE TWENTY-NINTH VOLUME.

PRINTED BY TAYLOR AND FRANCIS, LTD., RED LION COURT, FLEET STREET, E.C.4.